

10/19/31

DIALOG(R)File 347:JAPIO

(c) 2004 JPO & JAPIO. All rts. reserv

04818188 **Image available**
DISK CACHE DEVICE

PUB. NO.: 07-110788 [JP 7110788 A]
PUBLISHED: April 25, 1995 (19950425)
INVENTOR(s): YORIMITSU KEIICHI
 IWATANI SAWAO
APPLICANT(s): FUJITSU LTD [000522] (A Japanese Company or Corporation),
JP
 (Japan)
APPL. NO.: 05-256216 [JP 93256216]
FILED: October 14, 1993 (19931014)
INTL CLASS: [6] G06F-012/08
JAPIO CLASS: 45.2 (INFORMATION PROCESSING -- Memory Units)

ABSTRACT

PURPOSE: To perform an optimized disk cache operation suited to the successive application form of a user and a disk array by selectively setting a reliability considering mode, a performance considering mode or an average mode.

CONSTITUTION: Thus device is provided with a pair of disk control means 12 for reading and writing the data of a disk device 30 for access requests from a host device 10, a cache storage means 48 for storing a part at the data stored in the disk device 30, a cache control means 46 and a shared bus tar connecting the pair of the disk control means 12 and transmitting and receiving the information of the access requests or the like and the data. In this case, the cache control means 46 is provided with a reliability considering mode cache control means 72, a state considering mode cache control means 74-1 and an average mode cache control means 6-1. Then, cache control corresponding to an operating mode set by an operating mode setting means 70 at the time of starting the device is performed.

EG
242

This Page Blank (uspto)

特開平7-110788

(43) 公開日 平成7年(1995)4月25日

(51) Int.Cl.⁶

G 0 6 F 12/08

識別記号

3 2 0

庁内整理番号

7608-5B

F I

技術表示箇所

審査請求 未請求 請求項の数36 O L (全 42 頁)

(21) 出願番号

特願平5-256216

(22) 出願日

平成5年(1993)10月14日

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中1015番地

(72) 発明者 依光 圭一

神奈川県川崎市中原区上小田中1015番地

富士通株式会社内

(72) 発明者 岩谷 沢男

神奈川県川崎市中原区上小田中1015番地

富士通株式会社内

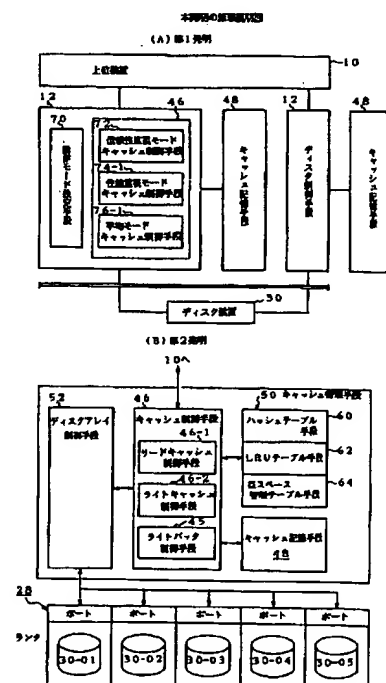
(74) 代理人 弁理士 竹内 進 (外1名)

(54) 【発明の名称】 ディスクキャッシュ装置

(57) 【要約】

【目的】 ユーザの運用形態とディスクアレイに適合して最適化されたディスクキャッシュ動作を提供する。

【構成】 2組のコントローラに各々キャッシュ機構を設け、各系統に同一データを格納する信頼性重視モード、各系統ごとに独立にデータを格納する性能重視モード、両者を折衷した平均モードのいずれかで動作可能とする。RAID 0, 1, 3, 5で動作するディスクアレイを対象としたキャッシュ処理を効率的に行なう。RAID 5の動作でもディスクリードを必要とせずパリティを計算できるRAID 3的なアクセスを可能とするようにキャッシュブロック上に有効データを準備する。

EG
10/2

【特許請求の範囲】

【請求項1】 ディスクキャッシュ装置に於いて、上位装置（10）からのアクセス要求に対しディスク装置（30）のデータ読み書きを制御する一対のディスク制御手段（12）と、前記一対のディスク制御手段（12）の各々に設けられ、前記ディスク装置（30）に格納されたデータの一部を記憶するキャッシュ記憶手段（48）と、前記一対のディスク制御手段（12）の各々に設けられ、前記上位装置（10）からのアクセス要求に対し予め準備された信頼性重視モードキャッシュ制御手段（72）、性能重視モードキャッシュ制御手段（74）、又は平均モードキャッシュ制御手段（76）にいずれか1つにより前記キャッシュ記憶手段（48）のデータを読み書きする一対のキャッシュ制御手段（46）と、前記一対のディスク制御手段（12）の間を結合して情報およびデータの送受を行う共用バス手段（18）と、装置の立ち上がり時に前記一対のキャッシュ制御手段（46）に対し予め準備された信頼性重視モード、性能重視モードまたは平均モードにいずれか1つの動作モードを設定する動作モード設定手段（70）と、を備えたことを特徴とするディスクキャッシュ装置。

【請求項2】 請求項1記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（12）に設けた前記信頼性重視モードキャッシュ制御手段（72）の各々は、前記一対のディスク制御手段（12）に設けたキャッシュ記憶手段（48）の各々に、同一のデータを記憶することを特徴とするディスクキャッシュ装置。

【請求項3】 請求項2記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（12）に設けた前記信頼性重視モードキャッシュ制御手段（72）の各々は、前記上位装置（10）からのリード要求に対し自己のキャッシュ記憶手段（48）に該当データがなかった場合は、前記ディスク装置（30）から該当データを読出して自己のキャッシュ記憶手段（48）にステージングして前記上位装置（10）にデータを転送すると共に、前記共通バス手段（18）を介して他のディスク制御手段（12）に転送して他方のキャッシュ記憶手段（48）に同じデータをステージングさせることを特徴とするディスクキャッシュ装置。

【請求項4】 請求項2記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（12）に設けた前記信頼性重視モードキャッシュ制御手段（72）の各々は、前記上位装置（10）からのライト要求に対し自己のキャッシュ記憶手段（48）にデータを書込むと共に、前記共通バス手段（18）を介して他のディスク制御手段（12）に転送して他方のキャッシュ記憶手段（48）にも同じデータを書込ませることを特徴とするディスクキャッシュ装置。

【請求項5】 請求項4記載のディスクキャッシュ装置に

於いて、前記一対のディスク制御手段（12）に設けた前記信頼性重視モードキャッシュ制御手段（72）の各々は、前記上位装置（10）からのライト要求の処理終了後に予め定めた書戻し条件が成立した際に、自己のキャッシュ記憶手段（48）からディスク装置（30）への書込みが済んでいないデータを抽出して前記ディスク装置（30）に書き戻すと共に、前記共用バス手段（18）を介して該書戻しデータのキャッシュ無効化を他方のディスク制御手段（12）に指示することを特徴とするディスクキャッシュ装置。

【請求項6】 請求項1記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（12）に設けた前記性能重視モードキャッシュ制御手段（74）の各々は、前記一対のキャッシュ記憶手段（48）の各々に、自己にアクセス要求のあったデータのみを個別に記憶することを特徴とするディスクキャッシュ装置。

【請求項7】 請求項6記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（12）に設けた前記性能重視モードキャッシュ制御手段（74）の各々は、前記上位装置（10）からのリード要求に対し自己のキャッシュ記憶手段（48）に該当データがなかった場合は、他方のディスク制御手段（12）に対し前記共用バス手段（18）を介して処理を依頼し、他のディスク制御手段（12）からのデータなしの応答を受けた際に前記ディスク装置から該当データを読出して自己のキャッシュ記憶手段（48）にステージングして前記上位装置（10）にデータを転送することを特徴とするディスクキャッシュ装置。

【請求項8】 請求項6記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（12）に設けた前記性能重視モードキャッシュ制御手段（74）の各々は、前記上位装置（10）からのライト要求に対し自己のキャッシュ記憶手段（48）にのみデータを書込むことを特徴とするディスクキャッシュ装置。

【請求項9】 請求項8記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（12）に設けた前記性能重視モードキャッシュ制御手段（74）の各々は、前記上位装置（10）からのライト要求の処理終了後に予め定めた書戻し条件が成立した際に、自己のキャッシュ記憶手段（48）からディスク装置（30）への書込みが済んでいないデータを抽出して前記ディスク装置（30）に書き戻すことを特徴とするディスクキャッシュ装置。

【請求項10】 請求項1記載のディスクキャッシュ装置に於いて、前記一対のディスク制御手段（12）に設けた前記平均モードキャッシュ制御手段（76）の各々は、前記上位装置（10）からのリード要求に伴う前記ディスク装置（30）からのステージングを自己のキャッシュ記憶手段（48）に対してのみ行い、上位装置（10）からのライト要求に伴うデータの書込みは、前

記一对のディスク制御手段（12）のキャッシュ記憶手段（48）の各々に行って同一データを記憶することを特徴とするディスクキャッシュ装置。

【請求項11】請求項10記載のディスクキャッシュ装置に於いて、前記一对のディスク制御手段（12）に設けた前記平均モードキャッシュ制御手段（76）の各々は、前記上位装置（10）からのリード要求に対し自己のキャッシュ記憶手段（48）に該当データがなかった場合は、他方のディスク制御手段（12）に対し前記共用バス手段（18）を介して処理を依頼し、該依頼後に前記ディスク手段から該当データを読出して自己のキャッシュ記憶手段（48）にステージングすることを特徴とするディスクキャッシュ装置。

【請求項12】請求項11記載のディスクキャッシュ装置に於いて、前記一对のディスク制御手段（12）に設けた前記平均モードキャッシュ制御手段（76）の各々は、他方のディスク制御手段（12）への処理依頼に対しデータなしの応答を受けていた際には、前記ディスク手段から該当データを読出して自己のキャッシュ記憶手段（48）にステージングした後に上位装置（10）に転送することを特徴とするディスクキャッシュ装置。

【請求項13】請求項10記載のディスクキャッシュ装置に於いて、前記一对のディスク制御手段（12）に設けた前記平均モードキャッシュ制御手段（76）の各々は、前記上位装置（10）からのライト要求に対し自己のキャッシュ記憶手段（48）にデータを書込むと共に、前記共通バス手段（18）を介して他のディスク制御手段（12）に転送して他方のキャッシュ記憶手段（48）にも同じデータを書込ませることを特徴とするディスクキャッシュ装置。

【請求項14】請求項13記載のディスクキャッシュ装置に於いて、前記一对のディスク制御手段（12）に設けた前記平均モードキャッシュ制御手段（76）の各々は、前記上位装置（10）からのライト要求の処理終了後に予め定めた書戻し条件が成立した際に、自己のキャッシュ記憶手段（48）からディスク装置（30）への書込みが済んでいないデータを抽出して前記ディスク装置（30）に書き戻すと共に、前記共用バス手段（18）を介して該書戻しデータのキャッシュ無効化を他方のディスク制御手段（12）に指示することを特徴とするディスクキャッシュ装置。

【請求項15】請求項1乃至14記載のディスクキャッシュ装置に於いて、前記一对のディスク制御手段（12）に設けたキャッシュ記憶手段（48）の各々を、不揮発性メモリ手段としたことを特徴とするディスクキャッシュ装置。

【請求項16】請求項1乃至5記載のディスクキャッシュ装置に於いて、前記一对のディスク制御手段（12）に設けたキャッシュ記憶手段（48）を揮発性メモリ手段としたことを特徴とするディスクキャッシュ装置。

【請求項17】請求項1乃至16記載のディスクキャッシュ装置に於いて、前記ディスク装置（30）をディスクアレイ装置としたことを特徴とするディスクキャッシュ装置。

【請求項18】ディスクキャッシュ装置に於いて、内蔵した記憶媒体に対しデータを読み書きする複数のディスク装置（30）を設けたディスクアレイ手段（28）と、

前記ディスクアレイ手段（28）に格納されたデータの一部を記憶するキャッシュ記憶手段（48）と、キャッシュ登録内容を示すハッシュテーブル手段（60）および最新に使用されたキャッシュブロックを先頭位置とするように有効データを含むキャッシュブロックを登録純にリンクするLRUテーブル手段（62）に基づいて前記キャッシュ記憶手段（48）の記憶状態を管理するキャッシュ管理手段（50）と、

前記上位装置（10）からリード要求を受けた際に前記キャッシュ管理手段（50）を参照して該当データが存在する場合は前記キャッシュ記憶手段（48）から読出して前記上位装置（10）に転送し、該当データが存在しない場合は、前記ディスクアレイ手段（28）からステージングした後に前記上位装置（10）に転送するリードキャッシュ制御手段（46-1）と、

前記上位装置（10）からライト要求を受けた際に、前記キャッシュ管理手段（50）に管理情報を登録すると共に該管理情報に従って前記キャッシュ記憶手段（48）にデータを書込むライトキャッシュ制御手段（46-2）と、

予め定めたライトバック条件が成立した際に、前記キャッシュ記憶手段（48）から前記ディスクアレイ手段（28）への記憶が済んでいないデータを抽出して書戻すライトバック制御手段（45）と、

前記リードキャッシュ制御手段（46-1）および前記ライトキャッシュ制御手段（46-2）による前記ディスクアレイ手段（28）のアクセス時に、予め設定されたディスクアレイの動作モードに従って1又は複数のディスク装置（30）のアクセスを制御するディスクアレイ制御手段（52）と、を備えたことを特徴とするディスクキャッシュ装置。

【請求項19】請求項18記載のディスクキャッシュ装置に於いて、前記ディスクアレイ手段（28）は、並列的に配置された入出力ポートの各々にディスク装置（30）を接続して1ランクを構成し、該ランク構成のディスク装置（30）群を複数ランク設けたことを特徴とするディスクキャッシュ装置。

【請求項20】請求項19記載のディスクキャッシュ装置に於いて、前記ディスクアレイ制御手段（52）は、RAID0に対応した第1動作モードの設定時は、特定のランクの中のディスク装置（30）を前記上位装置（10）で扱う論理デバイスに1対1に割当て、個々の

ディスク装置（30）に読み書きを行うことを特徴とするディスクキャッシュ装置。

【請求項21】請求項19記載のディスクキャッシュ装置に於いて、前記ディスクアレイ制御手段（52）は、RAID1に対応した第2動作モードの設定時は、特定のランクの中の2台のディスク装置（30）を1組として前記上位装置（10）で扱う論理デバイスに1対1に割当て、2つのディスク装置（30）に同一データを書込むと共にいずれか一方のディスク装置（30）からデータを読出すことを特徴とするディスクキャッシュ装置。

【請求項22】請求項19記載のディスクキャッシュ装置に於いて、前記ディスクアレイ制御手段（52）は、RAID3に対応した第3動作モードの設定時は、特定のランクを構成するn台のディスク装置（30）をデータ用とパリティ用とに固定的に割当て、書込要求時に、書込データをビット単位又はバイト単位に（n-1）分割すると共に分割単位ごとにパリティデータを計算し、前記（n-1）分割したデータおよび前記パリティデータを前記n台のディスク装置（30）に並列的に書込むことを特徴とするディスクキャッシュ装置。

【請求項23】請求項19記載のディスクキャッシュ装置に於いて、前記ディスクアレイ制御手段（52）は、RAID5に対応した第4動作モードの設定時は、特定のランクを構成するn台のディスク装置（30）を1組としてアクセスごとにパリティ位置が変化するように割当て、書込要求時に、書込データを少なくともディスク装置（30）のセクタ単位に分割すると共に該書込データの書込先ディスク装置（30）およびパリティ用ディスク装置から旧データおよび旧パリティデータを読み出して新パリティを計算し、前記書込データ及び新パリティデータを前記書込先ディスク装置（30）およびパリティ用ディスク装置に並列的に書込むことを特徴とするディスクキャッシュ装置。

【請求項24】請求項18記載のディスクキャッシュ装置に於いて、前記キャッシュ管理手段（50）は、前記キャッシュ記憶手段（48）の空き状態を管理する空きスペース管理テーブル手段（64）を備え、前記ライトキャッシュ制御手段（46-1）は前記上位装置（10）に基づくキャッシュ書込処理の完了報告時に前記空きスペース管理テーブル手段（60）を参照して空きスペースが予め定めた閾値以下の場合、ディスク書込済みのデータは無効化して前記キャッシュ記憶手段（48）から削除し、且つディスク書込みが済んでいないデータは前記ライトバック制御手段（45）による処理対象に組込むことを特徴とするディスクキャッシュ装置。

【請求項25】請求項24記載のディスクキャッシュ装置に於いて、前記空きスペース管理テーブル手段（64）は、前記キャッシュ記憶手段（48）上の連続するキャッシュブロックの空数をエントリとして先頭ブロックの

位置情報を格納したことを特徴とするディスクキャッシュ装置。

【請求項26】請求項18記載のディスクキャッシュ装置に於いて、前記リードキャッシュ制御手段（46-1）は、前記上位装置（10）からの要求データが前記キャッシュ記憶手段（48）の格納単位であるキャッシュブロック上に存在しない時、該キャッシュブロック上に存在する有効データの状態に応じて前記ディスクアレイ手段（28）から新たなデータのステージングを行って連続する一つのデータ領域を生成することを特徴とするディスクキャッシュ装置。

【請求項27】請求項26記載のディスクキャッシュ装置に於いて、前記リードキャッシュ制御手段（46-1）は、ステージング終了後に、前記要求データに続く有効データを前記ディスクアレイ手段（28）から前記キャッシュブロック上にプリフェッチすることを特徴とするディスクキャッシュ装置。

【請求項28】請求項27記載のディスクキャッシュ装置に於いて、RAID0に対応する第1動作モードの設定時又はRAID1に対応する第3動作モードの設定時に、前記リードキャッシュ制御手段（46-1）は、ステージング終了後に前記要求データに続く予め設定した量の有効データを前記ディスクアレイ手段（28）からプリフェッチすることを特徴とするディスクキャッシュ装置。

【請求項29】請求項27記載のディスクキャッシュ装置に於いて、RAID3に対応する第3動作モードの設定時又はRAID5に対応する第4動作モードの設定時に、前記リードキャッシュ制御手段（46-1）は、ステージング終了後に前記要求データに続く前記キャッシュブロックの最後まで有効データを前記ディスクアレイ手段（28）からプリフェッチすることを特徴とするディスクキャッシュ装置。

【請求項30】請求項27記載のディスクキャッシュ装置に於いて、RAID5に対応する第4動作モードの設定時に、前記リードキャッシュ制御手段（46-1）は、ステージング終了後に前記要求データを含むパリティグループの最後まで有効データを前記ディスクアレイ手段（28）からプリフェッチすることを特徴とするディスクキャッシュ装置。

【請求項31】請求項18記載のディスクキャッシュ装置に於いて、RAID5に対応する第4動作モードの設定時に、前記リードキャッシュ制御手段（46-1）は、オプション指定により前記キャッシュ記憶手段（48）にパリティデータを含めてステージングし、前記ライトバック手段（45）は、前記オプション指定を判別した際に前記キャッシュ記憶手段（48）のパリティデータを用いて新たなパリティを計算することを特徴とするディスクキャッシュ装置。

【請求項32】請求項18記載のディスクキャッシュ装

置に於いて、前記リードキャッシュ制御手段（４６－１）は、前記上位装置（１０）から大量シーケンシャルデータのリードが要求された場合、前記ＬＲＵテーブル手段（６２）の終端からの一定量となる既存のキャッシュブロックの一部を無効化すると共に、前記大量シーケンシャルデータを前記ＬＲＵテーブル手段（６２）の先頭に登録せずに終端側に登録することを特徴とするディスクキャッシュ装置。

【請求項３３】請求項１８記載のディスクキャッシュ装置に於いて、ＲＡＩＤ５に相当する第４動作モードの設定時に、前記ライトバック制御手段（４５）は、前記キャッシュブロックの中のライトバックの対象となった有効データをパリティグループ単位に抽出してパリティデータを算出した後に、抽出した有効データおよびパリティデータを前記ディスクアレイ手段（２８）の複数のディスク装置（３０）に並列的に書き込むことを特徴とするディスクキャッシュ装置。

【請求項３４】請求項１８記載のディスクキャッシュ装置に於いて、ＲＡＩＤ５に相当する第４動作モードの設定時に、前記ライトバック制御手段（４５）は、前記キャッシュブロックの中のライトバックの対象となった有効データのパリティグループに対する不足数が所定数以下の場合、該不足データを対応するディスク装置（３０）から読出した後にパリティグループ分の有効データおよび計算したパリティデータを前記ディスクアレイ手段（２８）の複数のディスク装置（３０）に並列的に書き込むことを特徴とするディスクキャッシュ装置。

【請求項３５】請求項３３又は３４記載のディスクキャッシュ装置に於いて、前記ライトバック手段（４５）は、所定のライトバック条件が成立した際に、前記ＬＲＵテーブル手段（６２）内の終端から所定数のキャッシュブロックをライトバック対象に組込むと共に、該対象範囲に含まれるキャッシュブロックをデータが連続するように並び替えることを特徴とするディスクキャッシュ装置。

【請求項３６】請求項１８記載のディスクキャッシュ装置に於いて、更に、装置の使用中に前記上位装置（１０）からのアクセス要求で使用されるデータブロックサイズを統計情報として収集し、該統計情報の平均値を次の装置立ち上げ時の前記キャッシュ記憶手段（４８）のキャッシュブロックサイズとして自動設定する統計処理手段を設けたことを特徴とするディスクキャッシュ装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、ホストコンピュータからの指示により磁気ディスクサブシステムのリードとライトを行うキャッシュメモリを備えたディスクキャッシュ装置に関し、更に、磁気ディスクサブシステムにディスクアレイを用いたディスクキャッシュ装置に関する。

【０００２】近年、コンピュータ本体の性能の向上に伴い、コンピュータ本体のデータ格納媒体である磁気ディスク装置を用いた入出力サブシステムの性能向上も望まれている。このため磁気ディスク装置よりもビットコストは高いが、高速アクセスが可能な半導体メモリをキャッシュメモリとして用いたディスクキャッシュ装置が登場した。

【０００３】すなわち、磁気ディスクサブシステムの一部のデータをキャッシュメモリ上に置き、キャッシュメモリをリードまたはライトすることでアクセスを高速化するという手法である。初期のディスクキャッシュ装置には、揮発性のメモリ素子が使用されており、電源障害等の発生でデータ消失に至る可能性があり、リード処理はキャッシュヒットになれば高速に処理できる。

【０００４】しかし、ライト処理は磁気ディスク装置とキャッシュメモリの両方の同時更新を実行しなければならないため、高速化できなかった。その後、不揮発性のメモリ素子を使用したディスクキャッシュ装置が出るに及び、ライト処理においてはライトデータの更新をキャッシュ上のみで行い、ライト処理の完了後にディスク装置に書き込むライトバック処理が可能となった。

【０００５】このライトバック処理の実現によりライト処理の高速化が行われた。しかし、メモリを不揮発性を維持するためのバッテリー容量とコストの問題等で不揮発性メモリと揮発性メモリを混在させ、コストと性能の最適化を図った装置が多い。しかし、不揮発性メモリと揮発性メモリを混在させたディスクキャッシュ装置では、ライト要求がリード要求よりも圧倒的に多いユーザ環境では、ライトバック処理を行う余裕がなく、ユーザに適應出来ないことが多く、より一般的で高速性と信頼性重視のある装置が望まれている。

【０００６】更に近年、ディスクアレイ装置を構成する複数の磁気ディスクを組合せ、ＲＡＩＤとして知られた動作形態をとるものが普及しだしている。これらＲＡＩＤ型ディスクアレイ装置に対してもディスクキャッシュ装置を適用し、最適化を図ることが望まれている。

【０００７】

【従来の技術】図４１は揮発性キャッシュメモリのみで構成した従来のディスクキャッシュ装置を示す。ホストコンピュータ１００からチャネル１０４－１、１０４－２を介してアクセスされる２台のコントローラ１０２－１、１０２－２が接続される。コントローラ１０２－１、１０２－２は、下位のドライブインタフェースによりアダプタ１０８－１、１０８－２を介してドライブとして磁気ディスク装置１１０－１、１１０－２、１１０－３を接続する。

【０００８】またコントローラ１０２－１、１０２－２は、下位のキャッシュインタフェースを介して共用される不揮発性のメモリ素子を用いたキャッシュメモリ１０６を接続している。図４２はキャッシュメモリに要求

データが存在した場合の動作を示す。ホストコンピュータ100から例えばコントローラ102-1にリード要求①があると、コントローラ102-1はキャッシュメモリ106に要求データが存在するか否かをハッシュテーブルを参照して調べるヒット判定②を行う。

【0009】すなわちキャッシュヒットの判定は、通常、磁気ディスク装置のアクセス単位であるシリンダ、トラック、ブロック等のハッシュパラメータからハッシュアドレス（ハッシュエントリ）を計算し、計算したハッシュアドレスで指定される要求データがハッシュテーブルに登録されているか否かをチェックする。要求データの存在を示すヒット判定が得られると、キャッシュメモリ106からデータ③を讀出してホストコンピュータ100に転送する。

【0010】一方、キャッシュメモリ106は要領的に制限があることから、LRUテーブル（Least Recent Use Table）を用いたキャッシュ管理が行われる。LRUテーブルは最新に使用されたキャッシュブロックをテーブルの先頭位置に登録し、最も古いキャッシュブロックがテーブルの最後に位置することになる。LRUテーブルの登録が一杯になると古いブロックから破棄し新たなブロックを登録する。また一定期間アクセスがないキャッシュブロックは、キャッシュメモリの使用効率を高めるため破棄する。このキャッシュブロックの登録を破棄してそのスペースを確保する処理をキャッシュ無効化という。キャッシュ無効化はハッシュテーブルの登録を削除することで実現できる。

【0011】図43はキャッシュメモリに要求データが存在しなかった場合の動作を示す。ホストコンピュータ100からのリード要求①に対するヒット判定②でデータが存在しないミスヒットになると、磁気ディスク装置110から要求データをリードしてキャッシュメモリ106に書込むステージング③を実行する。このステージング③は、磁気ディスク装置110の目的シリンダにヘッドを位置付けるシーク動作とシーク完了後のヘッド選択に基づく目的トラックへのオントラック動作等の機械的動作を伴うことから、キャッシュメモリ106のアクセスに比べ時間がかかる。

【0012】コントローラ102-1はキャッシュメモリ106のステージング③が完了すると、このデータ④をホストコンピュータ100に転送し、ハッシュテーブルにデータが存在すること（ヒット状態）を登録する。図44はライト要求に対する図41の装置の処理動作を示す。ホストコンピュータ100からライト指示①をコントローラ102-1が受けると、キャッシュメモリ106と磁気ディスク装置110の両方にデータを書込むライトスルー②を行う。

【0013】これはキャッシュメモリ106が揮発性であり、電源障害等でデータの消失、データ化け等発生する可能性があり、信頼性に対する考慮からコントローラ

102はデータをキャッシュメモリ106に書込むと同時に磁気ディスク装置110にも書込む。ライトスルーされたデータは次のリードでは必ずキャッシュヒットとなる。即ち、ライト処理に要する時間は、キャッシュなしの場合と同じ時間が必要でキャッシュ導入の効果はない。

【0014】ここで、図41の従来装置のライト処理は、磁気ディスク装置とキャッシュメモリの両方を同時更新を実行しなければならないため、高速化できなかった。そこで図45に示す揮発性キャッシュメモリと不揮発性キャッシュメモリを使用した装置が現われた。図45において、ホストコンピュータ200にチャネル204-1、204-2を介して接続された2台のコントローラ202-1、202-2は、下位のデバイスインタフェース側を全てのモジュール間での通信等を可能とする共有バス206を介して接続する。

【0015】共有バス206にはコントローラ202-1、202-2で共用される揮発性キャッシュメモリ206と不揮発性メモリ210が設けられる。更に共有バス206にコントローラ212-1、212-2とアダプタ214-1、214-2を介して複数の磁気ディスク装置216-1、216-2が接続される。ここで各系統ごとに複数のコントローラを設けて各々に固有の処理機能に持たせるマルチコントローラ構成とすることで、コントローラ単体の場合に比べて負荷を軽減させ、全体のスループットを向上させている。

【0016】揮発性キャッシュメモリ208は図41の従来装置で使用したものと同様である。不揮発性キャッシュメモリ210は、ディスク装置に対するデータ書込を高速化するために設けられ、揮発性キャッシュメモリ208に対する書込データ218と同じデータが書込まれる。不揮発性キャッシュメモリ210に書込まれたデータは、ライト処理の完了後にライトバックという処理で磁気ディスク装置に書込まれる。

【0017】図46はリード要求がミスヒットとなった時の図45の装置のステージングを示す。46において、ホストコンピュータ200からリード要求①を受けると、コントローラ202-1は不揮発性メモリ208のヒット判定②を行う。ミスヒットの場合は、磁気ディスク装置216から要求データをリードするステージング③を行い、キャッシュ管理テーブルに登録した後にホストコンピュータ200にデータ転送④を行う。

【0018】図47は図45の装置のライト処理を示す。46において、ホストコンピュータ200からライト指示①があると、コントローラ202-1は揮発性キャッシュメモリ208と不揮発性キャッシュメモリ210の両方にライトデータの書込み②を行い、キャッシュ管理テーブルを更新してホストコンピュータにライト完了を報告する。

【0019】不揮発性キャッシュメモリ210に書込さ

れたデータは、予め定めたライトバック処理条件が成立した時、例えば一定期間、リード又はライトされなかった場合、ライトバック対象データとして抽出され、磁気ディスク装置に書込むライトバック③が行われる。このライトバック③に伴いライトバック済みのデータにつき不揮発性キャッシュ210を無効化して破棄させる。一方、ライトバック済みデータは揮発性キャッシュ208では有効であり、リード要求に対してはヒット状態にある。

【0020】

【発明が解決しようとする課題】ところで、揮発性と不揮発性のキャッシュメモリを混在させた従来のディスクキャッシュ装置にあつては、揮発性キャッシュメモリの場合のネックであつたライト処理時間が、リード処理でキャッシュヒットした場合と同じ時間に改善されている。

【0021】しかしながら、不揮発性キャッシュメモリの容量が揮発性キャッシュメモリに比べて小さく、このためライトバックの回数が増加し、ライトバックのためにディスク側のコントローラと磁気ディスク装置の使用率が上昇し、システムのスループットが低下する問題がある。すなわち、不揮発性キャッシュメモリに格納できるデータ量が大きくとれないため、新規のライトデータを格納するには古いデータを磁気ディスク装置にライトバックして不揮発性キャッシュメモリを開放する必要が多く発生する。このため、ライトバック処理によりホストコンピュータからの新たな処理が実行出来ないことがある。

【0022】また不揮発性キャッシュメモリ自体に何らかの障害が発生した場合には、ライトバックができなくなり、ディスクキャッシュの機能が失われてしまう問題があつた。本発明は、このような従来の問題点に鑑みてなされたもので、磁気ディスクサブシステムの運用形態に応じた適合したキャッシュ機能を最大限に発揮できるディスクキャッシュ装置を提供することを目的とする。

【0023】一方、近年にあつては、計算機システムの外部記憶装置として、記録の不揮発性、大容量性、データ転送の高速性等の特長を持つ磁気ディスク装置、光ディスク装置等のディスク装置が広く用いられている。ディスク装置に対する要求は、高速データ転送、信頼性重視、大容量性、低価格である。これらの要求を満たすものとして、ディスクアレイ装置が注目されてきている。ディスクアレイ装置とは、小型ディスク装置を数台から数十台並べ、複数のディスク装置に分散してデータを記録して、並列的にアクセスする装置である。

【0024】ディスクアレイ装置で並列的に複数のディスク装置にデータ転送を行えば、一台のディスク装置の場合と比べて、ディスクの台数倍の高速データ転送が可能になる。また、データに加えて、パリティデータなどの冗長な情報を付け加えて記録しておくことで、ディス

ク装置の故障等を原因とするデータエラーの検出と訂正が可能となり、ディスク装置の内容を二重化して記録する方法と同程度の信頼性重視を、二重化より低価格で実現することができる。

【0025】従来、カルフォルニア大学バークレイ校のデビット・A・パターソン (David A. Patterson) らは、高速に大量のデータを多くのディスクにアクセスし、ディスク故障時におけるデータの冗長性を実現するディスクアレイ装置について、レベル1からレベル5までに分類付けを行つて評価した論文を発表している (ACMSIGMOD Conference, Chicago, Illinois, June 1-3, 1988 P109-P116)。

【0026】このデビット・A・パターソンらが提案したディスクアレイ装置を分類するレベル1～5は、RAID (Redundant Arrays of Inexpensive Disks) 1～5と略称される。RAID1～5を簡単に説明すると次のようになる。RAID0は、データの冗長性をもたないディスクアレイ装置であり、デビット・A・パターソンらの分類には含まれていないが、これを仮にRAID0と呼ぶ。

【0027】RAID1は、2台のディスク装置を1組として同一データを書込むミラーディスク装置であり、ディスク装置の利用効率が低い冗長性をもっており、簡単な制御で実現できるため、広く普及している。RAID2は、データをビットやバイト単位でストライピング (分割) し、それぞれのディスク装置に並列に読み書きを行う。ストライピングしたデータは全てのディスク装置で物理的に同じセクタに記録する。

【0028】データ用ディスク装置の他にハミングコードを記録するためのディスク装置を持ち、ハミングコードから故障したディスク装置を特定して、データを復元する。しかし、実用化されていない。RAID3は、データをビット又はバイト単位にストライピングしてパリティを計算し、ディスク装置に対しデータおよびパリティを並列的に書込む。RAID3は、大量のデータを連続して扱う場合には有効であるが、少量のデータをランダムにアクセスするトランザクション処理のような場合には、データ転送の高速性が生かせず、効率が低下する。

【0029】RAID4は、1つのデータをセクタ単位にストライピングして同じディスク装置に書込む。パリティは固定的に決めたディスク装置に格納している。データ書込みは、書込み前のデータとパリティを読み出してから新パリティを計算して書き込むため、1度の書込みについて、合計4回のアクセスが必要になる。また書込みの際に必ずパリティ用のディスク装置へのアクセスが起きるため、複数のディスク装置の書込みを同時に実行できない。このようにRAID4の定義は行われているが、メリットが少ないため現在のところ実用化の動きは少ない。

【0030】RAID5は、パリティ用のディスク装置を固定しないことで、並列的なリード、ライトを可能にしている。即ち、セクタごとにパリティの置かれるディスク装置が異なっている。パリティディスクが重複しなければ異なるディスク装置にセクタデータを並列的に書込むことができる。このようにRAID5は非同期に複数のディスク装置にアクセスしてリード又はライトを実行できるため、少量データをランダムにアクセスするランザクション処理に向いている。

【0031】このようなRAID型で分類されるディスクアレイ装置は、システムの運用形態に応じてRAIDの動作形態が選択されているが、最近では、1つのディスクアレイ装置で、RAID0、1、3および5に動作モードを異なるディスク装置に割当ててホストコンピュータが論理デバイスとして選択指定できるようにしたものが出されている。

【0032】このようなRAID型のディスクアレイ装置においても、ディスクキャッシュ装置を使用することで処理性能をより一層向上させることができるが、RAIDの動作形態に適したディスクキャッシュの機能を最適化する必要がある。従って、本発明の他の目的は、RAID型で動作形態が分類されたディスクアレイ装置を対象にディスクキャッシュ機能の最適化を図ったディスクキャッシュ装置を提供することを目的とする。

【0033】

【課題を解決するための手段】図1は本発明の原理説明図である。

【第1発明】まず第1発明のディスクキャッシュ装置は、図1(A)に示すように、上位装置10からのアクセス要求に対しディスク装置30のデータを読み書きする一対のディスク制御手段12と、ディスク制御手段12の各々に設けられ、ディスク装置30に格納されたデータの一部を記憶するキャッシュ記憶手段48と、一対のディスク制御手段12の各々に設けられたキャッシュ制御手段46と、一対のディスク制御手段12の間を結合してアクセス要求などの情報およびデータの送受を行う共用バス手段18とを備える。

【0034】ここでキャッシュ制御手段46には、信頼性重視モードキャッシュ制御手段72、性能重視モードキャッシュ制御手段74、又は平均モードキャッシュ制御手段76が設けられ、動作モード設定手段70により装置の立上り時に設定された動作モードに従ったキャッシュ制御が行われる。

I) 信頼性重視モード

動作モード設定手段70で信頼性重視モードを設定した場合、一対のディスク制御手段12に設けた信頼性重視モードキャッシュ制御手段72の各々は、一対のディスク制御手段12に設けたキャッシュ記憶手段48の各々に、同一のデータを記憶する。

【0035】具体的には、上位装置10からのリード要

求に対し自己のキャッシュ記憶手段48に該当データがなかった場合（ミスヒット）は、ディスク装置30から該当データを読み出してステージングして上位装置10にデータを転送し、同時に共用バス手段18を介して他のディスク制御手段12に転送して他方のキャッシュ記憶手段48-2に同じデータをステージングさせる。

【0036】また上位装置10からのライト要求に対しては、自己のキャッシュ記憶手段48にデータを書込むと共に、共用バス手段18を介して他のディスク制御手段12に転送して他方のキャッシュ記憶手段（48）にも同じデータを書込ませる。更にライト要求の処理終了後のライトバック処理として、ライトバック条件の成立時に、自己のキャッシュ記憶手段48からディスク装置30への書込みが済んでいないデータ（以下「ダーティデータ」という）を抽出してディスク装置30に書き戻すと共に、共用バス手段18を介して書き戻済みデータのキャッシュ無効化を他方のディスク制御手段12に指示する。

II) 性能重視モード

動作モード設定手段70で性能重視モードを設定した場合は、一対のディスク制御手段12に設けた性能重視モードキャッシュ制御手段74の各々は、一対のディスク制御手段12に設けたキャッシュ記憶手段48の各々に、自己にアクセス要求のあったデータのみを個別に記憶する。

【0037】具体的には、上位装置10からのリード要求に対し自己のキャッシュ記憶手段48に該当データがなかった場合は、他方のディスク制御手段12に対し共用バス手段18を介して処理を依頼し、他のディスク制御手段12からのデータなしの応答を受けた際にディスク装置30から該当データを読み出して自己のキャッシュ記憶手段48にステージングして上位装置10にデータを転送する。

【0038】また上位装置10からのライト要求に対しては、自己のキャッシュ記憶手段48にのみデータを書込む。ライトバック処理は、書き戻条件が成立した際に、自己のキャッシュ記憶手段48からディスク装置30へ書込みが済んでいないダーティデータを抽出してディスク装置30に書き戻す。

III) 平均モード

動作モード設定手段70で平均モードを設定した場合、一対のディスク制御手段12に設けた平均モードキャッシュ制御手段76の各々は、一対のディスク制御手段12の各々は、上位装置10からのリード要求に伴うディスク装置30からのステージングは自己のキャッシュ記憶手段48に対してのみ行い、上位装置10からのライト要求に伴うデータの書込みは、一対のディスク制御手段12のキャッシュ記憶手段48の各々に行って同一データを記憶する混在型とする。

【0039】具体的には、上位装置10からのリード要

求に対し自己のキャッシュ記憶手段48に該当データがなかった場合は、他方のディスク制御手段12に対し共用バス手段18を介して処理を依頼し、依頼後にディスク装置30から該当データを読出して自己のキャッシュ記憶手段48にステージングする。ここで他方のディスク制御手段12への処理依頼に対しデータなしの応答を受けていた際には、ディスク装置30から該当データを読出して自己のキャッシュ記憶手段48にステージングした後上位装置10に転送する。

【0040】また上位装置10からのライト要求に対しては、自己のキャッシュ記憶手段48にデータを書込むと共に、共通バス手段18を介して他のディスク制御手段12に転送して他方のキャッシュ記憶手段48にも同じデータを書込ませる。この点は信頼性モードと同じである。更にライトバック処理は、書戻し条件が成立した際に、自己のキャッシュ記憶手段48からディスク装置30への書込みが済んでいないダーティデータを抽出してディスク装置30に書き戻すと共に、共用バス手段18を介して書戻しデータのキャッシュ無効化を他方のディスク制御手段12に指示する。これは性能重視モードと同じである。

【0041】IV) メモリ構成

更に、信頼性重視モード、性能重視モード及び平均モードのいずれについても、一対のディスク制御手段12に設けたキャッシュ記憶手段48の各々を、不揮発性メモリとする。但し、信頼性重視モードについては、両方とも揮発性メモリでもよい。またディスク装置30をディスクアレイ装置としてもよい。

【第2発明】図1(B)はディスクアレイ装置のRAID形態に適合した第2発明のディスクキャッシュ装置の原理説明図である。

【0042】このディスクキャッシュ装置は、複数のディスク装置30を設けたディスクアレイ手段28を備える。ディスクキャッシュ機構としては、ディスクアレイ手段28に格納されたデータの一部を記憶して上位装置10からのアクセス要求に対しデータを読み書きするキャッシュ記憶手段48を持ち、キャッシュ管理手段50のハッシュテーブル手段60およびLRUテーブル手段62に基づいてキャッシュ記憶手段48の記憶状態を管理する。

【0043】またリードキャッシュ制御手段46-1とライトキャッシュ制御手段46-2が設けられる。リードキャッシュ制御手段46-1は、上位装置10からリード要求を受けた際に、キャッシュ管理手段50のハッシュテーブル手段60を参照し、要求データが存在する場合は、キャッシュ記憶手段48から読出して上位装置10に転送する。要求データが存在しない場合は、ディスクアレイ手段28からステージングした後上位装置10に転送する。

【0044】ライトキャッシュ制御手段46-2は、上

位装置10からライト要求を受けた際に、キャッシュ管理手段50のハッシュテーブル手段60に書込を示す情報を登録すると共に、キャッシュ記憶手段48にデータを書込み、書込み済みのキャッシュブロックをLRUテーブル手段62の先頭にリンクする。またライトバック制御手段が設けられ、予め定めたライトバック条件が成立した際に、キャッシュ記憶手段48からディスクアレイ手段28への記憶が済んでいないデータを抽出して書戻す。

【0045】更にディスクアレイ制御手段50が設けられ、リードキャッシュ制御手段46-1および前記ライトキャッシュ制御手段46-2によるディスクアレイ手段28のアクセス時に、予め設定されたディスクアレイのRAID動作モードに従って1又は複数のディスク装置30のアクセスを制御する。

I) RAID動作モード

ディスクアレイ手段28は、並列的に配置された入出力ポートの各々にディスク装置30を接続して1ランクを構成し、この1ランク構成のディスク装置群を複数ランク設けている。ディスクアレイ制御手段52は、RAID0、RAID1、RAID3又はRAID5に対応した第1～第4動作モードを設定する。

【0046】RAID0に対応した第1動作モードの設定時は、特定のランクの中のディスク装置30を上位装置10で扱う論理デバイスに1対1に割当て、個々のディスク装置30に読み書きを行う。RAID1に対応した第2動作モードの設定時は、特定のランクの中の2台のディスク装置30を1組として上位装置10で扱う論理デバイスに1対1に割当て、2つのディスク装置30に同一データを書込むと共にいずれか一方のディスク装置30からデータを読出すミラーディスクとする。

【0047】RAID3に対応した第3動作モードの設定時は、特定のランクを構成するn台のディスク装置30をデータ用とパリティ用とに固定的に割当て、上位装置10からの書込要求時に、書込データをビット単位又はバイト単位に(n-1)分割すると共に分割単位ごとにパリティデータを計算し、(n-1)分割したデータおよびパリティデータをn台のディスク装置30に並列的に書込む。

【0048】RAID5に対応した第4動作モードの設定時は、特定のランクを構成するn台のディスク装置30を1組としてアクセスごとにパリティ位置が変化するように割当て、上位装置10からの書込要求時に、書込データを少なくともセクタ単位に分割すると共に、書込データの書込先ディスク装置30およびパリティ用ディスク装置から旧データおよび旧パリティデータを読出して新パリティを計算し、前記書込データ及び新パリティデータを前記書込先ディスク装置30およびパリティ用ディスク装置に並列的に書込む。

II) キャッシュメモリの空スペースの確保

キャッシュ管理手段50は、キャッシュ記憶手段48の空き状態を管理する空スペース管理テーブル手段60を備える。ライトキャッシュ制御手段46-1は上位装置10に基づくキャッシュ書込処理の完了報告時に空スペース管理テーブル手段64を参照し、空スペースが予め定めた閾値以下の場合、ディスク書込済みのデータは無効化してキャッシュ記憶手段48から削除し、且つディスク書込みが済んでいないデータはライトバック手段45による処理動作を準備する。これにより常に一定量の空スペースがキャッシュ記憶手段48に確保され、ライト動作が迅速にできる。

【0049】空スペース管理テーブル手段64には、キャッシュ記憶手段48上の連続するキャッシュブロックの空ブロック数をエントリ（インデックス）として先頭ブロックの位置情報を格納する。

III) ステージング動作]

リードキャッシュ制御手段46-1は、上位装置10からのリード要求データがキャッシュ記憶手段48の格納単位であるキャッシュブロック上に存在しない時、キャッシュブロック上に存在する有効データの状態に応じてディスクアレイ手段28から新たなデータのステージングを行って連続する一つのデータ領域を生成する。

【0050】ステージング終了後に、リード要求データに続く有効データをディスクアレイ手段28からキャッシュブロック上にプリフェッチする。このプリフェッチするデータ量は、RAID1に対応する第3動作モードの設定時は、予め設定した量の有効データとする。RAID3に対応する第3動作モードの設定時又はRAID5に対応する第4動作モードの設定時には、キャッシュブロックの最後まで有効データとする。

【0051】RAID5に対応する第4動作モードの設定時には、リード要求データを含むパリティグループの最後まで有効データとしてもよい。これはキャッシュブロックサイズとパリティグループのブロックサイズが一致する場合である。RAID5に対応する第4動作モードの設定時に、リードキャッシュ制御手段46-1は、オプション指定によりキャッシュ記憶手段48にパリティデータもステージングし、この場合、ライトバック手段45は、キャッシュ記憶手段48のパリティデータを用いて新たなパリティを計算する。これによりライトバックの際の旧データおよび旧パリティのディスク読出を不要にできる。

IV) シーケンシャルデータのLRU管理

リードキャッシュ制御手段46-1は、上位装置10から大量シーケンシャルデータのリード要求に対しステージングを行った場合、キャッシュ記憶手段48のLRUテーブル手段62を参照して末尾から所定量のキャッシュブロックとなる記憶データの一部を無効化すると共に、大量シーケンシャルデータをLRUテーブル手段62の先頭に登録せずに終端側に登録する。これによって

比較的短期間で大量シーケンシャルデータをキャッシュ記憶手段から追出すことができる。

V) ライトバック処理]

RAID5に相当する第4動作モードの設定時に、ライトバック手段45は、キャッシュブロックの中のライトバックの対象となった有効データをパリティグループ単位に抽出してパリティデータを算出した後に、抽出した有効データおよびパリティデータをディスクアレイ手段28の複数のディスク装置30に並列的に書込む。即ち、RAID5であってもRAID3と同様に動作させてライトバックを効率良く行う。

【0052】またライトバック手段45は、キャッシュブロックの中のライトバックの対象となった有効データのパリティグループに対する不足数が所定数以下の場合、例えば1ブロック以下の場合、不足データに対応するディスク装置30から読出した後にパリティグループ分の有効データおよび計算したパリティデータをディスクアレイ手段28の複数のディスク装置30に並列的に書込むことでRAID3的な動作を可能とする。

【0053】ライトバック手段45は、所定のライトバック条件が成立した際に、LRUテーブル手段62内の終端から所定数のキャッシュブロックをライトバック対象に指定すると共に、この範囲に含まれるキャッシュブロックをデータが連続するように並び替え、RAID3的なライトバックを可能とする。

VI) キャッシュブロックサイズの自動設定]

更に本発明は、装置の使用中に上位装置10からのアクセス要求で使用するデータブロックサイズを統計情報として収集し、この統計情報の平均値を次の装置立ち上げ時のキャッシュ記憶手段48のキャッシュブロックサイズとして自動設定する統計処理手段を設ける。

【0054】

【作用】まず図1(A)の第1発明の作用を説明する。第1発明のディスクキャッシュ装置で信頼性重視モードを設定した場合には、両系統のキャッシュメモリ内のデータを常に同一とするようにキャッシュ動作が行われる。即ち二重化されたキャッシュメモリを独立に使用するのではなく、片側データをバックアップ用と考えている。

【0055】この場合の長所は、データが常に二重化されており片側データの喪失時データが保障され高い信頼性が得られる。また短所としては、キャッシュメモリの半分がバックアップ用であり、使用効率が低くなる。また第1発明のディスクキャッシュ装置で性能重視モードを設定した場合には、各キャッシュメモリのデータは独立であり、信頼性モードの最大2倍の容量のキャッシュメモリとして使用できる。

【0056】この場合の長所は、両系統のキャッシュメモリが独立に動作し、また他方からの参照更新が可能であるためキャッシュ容量が増大する。短所としては、メ

モリ障害、電源障害でデータ消失、データ化け等の障害が発生する確立が高くなる。また他系統にキャッシュ処理を依頼したヒットとなった時は、リモート側キャッシュに対してのリードとなるので性能が低下する。

【0057】更に第1発明のディスクキャッシュ装置で平均的モードを設定した場合は、基本的には信頼性重視モードと性能重視モードの混在型となる。すなわち、リード要求のミスヒット時のステージングは自己のキャッシュメモリに対してのみ行い、またデータのライトは両方系統のキャッシュメモリに対して行う。これによりライト処理の少ない環境ではキャッシュ容量は性能重視モードと同様に2倍となり、またライト処理が多く発行される環境では信頼性重視モードと同様に障害に対して強い形になる。この場合の長所と短所は、使用環境、即ちリードとライトの発行比率に応じていずれかの信頼性重視モードと性能重視モードの長所と短所が強く現れることになる。

【0058】次に図1(B)のディスクアレイを対象としたディスクキャッシュ装置の作用を説明する。まず第2発明にあつては、キャッシュメモリの管理に使用する空スペース管理テーブルを備える。すなわちキャッシュメモリ上にないデータに対してのステージング又はライトが指示された場合、コントローラはキャッシュメモリ上の未使用の空間から新たなキャッシュブロックを割当て、ディスク装置からのステージングによるデータ、あるいはホストコンピュータからのライトデータを受領する。

【0059】この場合、キャッシュメモリ上に一定量の空き空間が残るように空スペース管理テーブルを用いて管理される。すなわち残り容量が規定値以下になるような状況が発生すると、キャッシュメモリ上の有効データの追い出しがスケジュールされる。即ち、ディスク書込が済んだデータはハッシュエントリで指定されるLRUテーブルの登録を抹消する無効化を行う。またディスク書込みが済んでいないダーティデータは、ディスクへの書き戻しがスケジュールされる。これにより新たな上位装置からのリード又はライト要求に対し常にキャッシュメモリ上に空き空間が用意され、直ちに処理を開始できる。

【0060】次に第2発明にあつては、リード要求に対しキャッシュメモリがミスヒットとなった時のディスク装置からのステージングを、キャッシュメモリを使用したことによる効果が最大に発揮できるように行う。これは要求されたホストブロック分のデータをディスク装置からキャッシュブロック上にステージングした後に、後続するホストブロックに対応するデータをディスク装置からプリフェッチすることで実現される。

【0061】このようにステージングが済んだホストブロックに引き続くブロックに対して先読み動作を行うことで、ホストコンピュータからの逐次リード要求に対し

てキャッシュ効果を最大限にする処理ができる。この場合のプリフェッチ量は、RAID0、1、3、5で異なる。まずRAID3は1つのホストブロックが所定バイト数単位に各ディスク装置に分散して格納されており、ステージングが高速に行われること及び大容量データが要求される場合が多いことを考慮し、ステージングされた有効ブロックを含むキャッシュブロックの最後までホストブロックに対応するデータをプリフェッチする。

【0062】RAID5は、基本的にはリード転送速度はディスク装置の速度以上は出ず、ステージング速度がRAID3より遅いこと、また小容量データのリード処理が中心になることから、ステージングが済んだブロックに続く次のパリティグループ（パリティ用ディスク装置を同一とするグループ）までとする。ここで、キャッシュブロックサイズとパリティグループのブロックサイズが同一となるようにストライピングを行えば、キャッシュブロックの最後迄がプリフェッチ量となり、RAID3的な高速ステージングが可能となる。

【0063】更にプリフェッチ後のキャッシュブロック上で、連続する有効ブロックに空きがあつて分散している場合には、この空ブロックに内部的なステージングを実行し、キャッシュブロック上の有効データを連続ブロック領域とし、データ管理を容易にする。また第2発明のディスクキャッシュ装置は、大量シーケンシャルデータのステージングに対し、その追い出しを可能ながざり早めるようにキャッシュメモリを管理する。

【0064】一般にディスクキャッシュ装置は同一データを頻繁にリードすると効果が出る。ホストコンピュータがキャッシュメモリ上に存在しない大量データのリード要求を行ったとき、ディスク装置からデータをステージングしホストに転送する。このデータは一度しか参照されないにも関わらずそのままキャッシュメモリ上に残しておくと、キャッシュスペースが有効に利用されないことになる。しかし直ちに無効化することも出来ない。

【0065】そこで、ホストコンピュータが一定量以上のデータ、例えばキャッシュブロック100分相当等を要求した時、ステージング後、本来はLRUテーブルの先頭位置に登録すべきところを、第2発明は、LRUテーブルの終端の近傍に登録する。これにより比較的短い時間で、大量シーケンシャルデータは擬似的に最も使用されていない古いデータとなり、キャッシュ無効化により追い出すことができる。

【0066】また第2発明は、オプション指定によりRAID5でディスクアレイからパリティデータを含めてキャッシュメモリにステージングすることができる。このようなパリティデータを含めたステージングにより、ライトバック処理でディスクアレイからの旧データおよび旧パリティデータのリードが不要となり、RAID5のもつライトペナルティを軽減させることができる。

【0067】具体的には、パリティのステージングを行

わない場合、ディスク装置の2.5回転分の時間を必要とするが、パリティをステージングしていた場合には、1.5回転分の時間で済む。但し、キャッシュメモリの容量はパリティをステージングした分だけ減少する。さらに第2発明のディスクキャッシュ装置は、RAID5の動作状態で、ライト処理後に行なうディスクアレイに対するライトバック処理を最適化する。

【0068】RAID5では、パリティグループを構成する全てのディスク装置に跨がって連続するデータがキャッシュブロック上に存在する場合は、RAID3的に連続データからパリティを計算し、データおよびパリティの同時転送で並列的に書き込むことができる。一方、パリティグループを構成する一部のディスク装置のデータしかキャッシュブロック上に存在しない場合は、旧パリティと旧データをリードしてパリティを計算した後にデータとパリティを並列的に書き込む通常のリードモディファイライトを行う。

【0069】この場合、キャッシュブロック上で連続するデータの空きが所定値以下、例えば1ディスク装置分であった場合には、空部分に有効データをステージングして連続データとした後に、RAID3的な同時更新のライトバックを行う。更に、キャッシュメモリの空きスペースが規定値以下となってライトバック処理がスケジュールされた場合、LRUテーブルの末尾から一定範囲のキャッシュブロックをライトバック対象に指定する。このライトバック対象として指定されたキャッシュブロックに対し、ホストブロックが連続するように並び替えを行い、RAID3的な高速のライトバック処理を可能とする。

【0070】更に第2発明のディスクキャッシュ装置は、ユーザの運用形態に応じて自動的にキャッシュブロック・サイズを設定することができる。一般にユーザシステムにおいてユーザが必ず設定を必要とするパラメータは、ホストブロックサイズ（ロジカルブロックサイズ）である。キャッシュブロックサイズは、ホストブロックサイズが複数集まったサイズとして設定される。第2発明は、最初のシステム立ち上げ時は、デフォルトのキャッシュブロックサイズで初期動作するが、次の立ち上げ時には、ユーザの使用環境に応じて自動的にキャッシュブロックサイズを変更する。

【0071】即ち、ユーザ使用中にキャッシュシステムをアクセスしたホストブロックサイズの統計を取り、平均的なホストブロックサイズを求めている。そして次の電源投入に伴う立ち上げ時に、前回の運用で求めた新規なキャッシュブロックサイズでキャッシュシステムを動作させ、キャッシュブロック・サイズをユーザの判断を必要とすることなく最適化する。

【0072】

【実施例】

<目次>

1. システムのハードウェア構成
2. 第1発明の全体機能
3. 信頼性重視モードのディスクキャッシュ制御
4. 容量重視モードのキャッシュ制御
5. 平均モードのキャッシュ制御
6. 第2発明の全体機能
7. キャッシュ空スペースの確保
8. RAID動作形態の制御
9. ステージング処理の高速化
10. 大量シーケンシャルデータの追出し促進
11. パリティデータのステージング
12. ライトバックデータの並び替えによる最適化
13. ライトバック処理の最適化
14. 統計的処理によるキャッシュブロックサイズの自動設定

1. システムのハードウェア構成

図2は本発明のディスクキャッシュ装置が適用される磁気ディスク装置を用いた入出力サブシステムのハードウェア構成を示す。

【0073】上位装置としてのホストコンピュータ10には少なくとも2つのチャネル装置14-1, 14-2が設けられ、チャネル14-1, 14-2に対し、チャネルインタフェース16を介して2台のコントローラ12-1, 12-2を接続している。この実施例で、チャネルインタフェース16としてはSCSIを使用している。勿論、MBCインタフェース（ブロック・マルチプレクサ・チャネルインタフェース）を使用してもよい。

【0074】コントローラ12-1, 12-2はディスク制御手段としての機能を有し、キャッシュ機構を内蔵している。コントローラ12-1, 12-2のデバイス側は共用バス18-1, 18-2に接続される。共用バス18-1, 18-2はブリッジ回路部20で結合され、コントローラ12-1, 12-2間でメッセージ、コマンド等の情報およびデータのやり取りを行うことができる。更に共用バス18-1, 18-2はサブコントローラ22-1, 22-2が接続され、コントローラ12-1, 12-2の処理機能を分散させて負荷の低減を図っている。

【0075】共用バス18-1, 18-2にはアダプタ24-1~24-6, 26-1~26-6のそれぞれを介して、ディスクアレイ28に設けている複数のディスク装置30-00~30-35が接続される。この実施例でディスクアレイ28はコントローラ12-1, 12-2より並列アクセスを受ける6つのポートP0~P5で並列ディスク群を構成し、この並列ディスク群をランクR0~R3で示す4ランク分設けている。

【0076】具体的には、ランクR0はポートP0~P5に対応した6台のディスク装置30-00~30-05で構成され、ランクR1はポートP0~P5に対応したディスク装置30-10~30-15で構成され、ラ

ンクR2はポートP0～P5に対応したディスク装置30-20～30-25で構成され、更にランクR3はポートP0～P5に対応したディスク装置30-30～30-35で構成される。

【0077】このようなディスクアレイを構成するディスク装置の位置はランク番号Rとポート番号Pアドレスで定義する。例えば磁気ディスク装置30-00は(R0, P0)で表わすことができる。図3は図2のコントローラ12-1側のハードウェア構成を示す。コントローラ12-1内にはCPU32が設けられ、CPU32の内部バス44にROM34、DRAM36、SCSI回路部40とのやり取りを行う上位インタフェース部38、共用バス18-1とのやり取りを行うバスインタフェース部42が設けられる。

【0078】更にディスクキャッシュ機構を実現するためキャッシュ制御部46とキャッシュメモリ48が設けられる。なお、キャッシュ制御部46を独立に設けて、CPU32の機能として実現してもよい。またキャッシュ制御部46においてキャッシュメモリ48の管理に使用されるキャッシュ管理テーブル、具体的にはキャッシュエントリテーブル、LRUテーブルおよび空きスペース管理テーブルはDRAM36に展開されている。

2. 第1発明の全体機能

図4は図2のハードウェア構成を対象とした第1発明の処理機能を示す。尚、説明を簡単にするため、ディスクアレイ側については1台のディスク装置30を代表して示し、共用バスについても1本の共用バス18で示している。またキャッシュメモリについてはコントローラから取出して示している。

【0079】コントローラ12-1、12-2には動作モード設定部70-1、70-2、信頼性重視モードキャッシュ制御部72-1、72-2、性能重視モードキャッシュ制御部74-1、74-2、および平均モードキャッシュ制御部76-1、76-2が設けられている。キャッシュメモリとしては不揮発性キャッシュメモリ48-1、48-2が使用される。

【0080】ここでコントローラ12-1側を例にとると、動作モード設定部70-1は予め定めたホストコンピュータ10からのコマンド指示あるいはコントローラ12-1に予め設けられたピン設定などにより、信頼性重視モード、性能重視モードまたは平均モードのいずれか1つを設定する。この動作モード設定部70-1による設定に基づき、信頼性重視モードキャッシュ制御部72-1、72-2、性能重視モードキャッシュ制御部74-1、74-2、または平均モードキャッシュ制御部76-1、76-2のいずれか1つの動作を有効とする。

【0081】ここでホストコンピュータ10に対し2台のコントローラ12-1、12-2が設けられていることから、ホストコンピュータ10からのアクセス要求を

受けたコントローラ側をローカル側といい、このときアクセス要求を受けていない側をリモート側という。これはコントローラ12-1、12-2において相対的なものであり、コントローラ12-1に対しホストコンピュータ10からアクセス要求が行われたときはコントローラ12-1がローカル側、コントローラ12-2がリモート側となる。

【0082】逆にコントローラ12-2に対しホストコンピュータ10からアクセス要求が行われたときはコントローラ12-2がローカル側、コントローラ12-1がリモート側となる。

3. 信頼性重視モードのディスクキャッシュ制御

信頼性重視モードにあつては、コントローラ12-1、12-2に設けたキャッシュメモリは図4に示したように両方とも不揮発性キャッシュメモリ48-1、48-2とする。

【0083】尚、信頼性重視モードにあつては、キャッシュメモリを両方とも揮発性としてもよい。この場合、後の説明で明らかにする性能重視モードおよび平均モードにあつては、両方とも不揮発性キャッシュメモリとする必要があるため、電源バックアップの有無により不揮発性となるか揮発性と不揮発性が切替可能なメモリを使用する必要がある。

【0084】信頼性重視モードにあつては、コントローラ12-1、12-2の各不揮発性キャッシュメモリ48-1、48-2のデータが同一であるようにキャッシュ動作が行われる。即ち、2系統のキャッシュメモリを独立に使用するのではなく、同一データを格納することで二重化し、片側のデータ送出時にあつても、キャッシュデータを保証して信頼性を高める。

【0085】リード要求に対するミスヒット時のステージングはディスク装置30から両方の不揮発性キャッシュメモリ48-1、48-2に対し行う。またライト要求に対するライトデータの書込みも両方の不揮発性48-1、48-2に対し行う。更に、ライト処理終了後のディスク装置30に対するライトバックについては、ライト要求を受けたローカル側のキャッシュメモリから行う。

【0086】図5は信頼性重視モードにおけるキャッシュ動作を示したフローチャートである。まずステップS1にあつては、ホストコンピュータ10からのアクセス要求の有無を監視しており、アクセス要求があるとステップS2に進み、リード要求かライト要求かのアクセスモードを判別する。リード要求であれば、ステップS3で自己の不揮発性キャッシュメモリ即ちローカル側キャッシュメモリに要求データが存在するか否かのヒット判定を行う。

【0087】キャッシュヒットが得られれば、ステップS4で要求データを読み出してホストコンピュータ10にデータ転送する。キャッシュミスヒットの場合にはデ

ディスク装置30からのステージングを行うことになるが、ステージングに先立ち、リモート側のコントローラ12-1にステージングを通知する。このステップS5における処理は図7に示される。即ち、ホストコンピュータ10からのリード要求を受けたコントローラ12-1が不揮発性キャッシュメモリ48-1のミスヒットを判定すると、ディスク装置30からのステージングに先立ってローカル側のコントローラ12-2に対し共用バス18を経由してメッセージを送る。

【0088】このメッセージはキャッシュミスヒットとなったため、ステージングによりディスク装置30からのデータをリモート側の不揮発性キャッシュメモリ48-2にも転送することを通知する。このローカル側のコントローラ12-1からのメッセージを受けてリモート側のコントローラ12-2は自己のキャッシュ管理テーブルを更新し、不揮発性キャッシュメモリ48-2内にステージング用の空間を割り当て、スタンバイ状態となったことをローカル側のコントローラ12-1に送り返す。

【0089】再び図5を参照するに、ステップS5でリモート側へのステージングの通知が済むと、リモート側からのレディ応答を待って、ステップS6で要求データをディスク装置30から読み出し、ローカル側の不揮発性キャッシュメモリ48-1に格納すると共に、共用バス18を介してリモート側の不揮発性キャッシュメモリ48-2に転送し、それぞれ書き込むステージングを行う。

【0090】ローカル側の不揮発性キャッシュメモリ48-1に対するステージングが済むと、ステージングにより得られた要求データ④をホストコンピュータ10に転送し、ステップS8でリード終了応答を行って一連のリード要求に対する処理を終了する。次にステップS2でライト要求を判別した場合には、ライト要求を受けたローカル側のコントローラ12-1がリモート側のコントローラ12-2へ共用バス18を介してライトデータの転送を通知する。

【0091】例えば図8に示すように、ホストコンピュータ10よりライト要求①をコントローラ12-1が受けると、共用バス18を経由してリモート側のコントローラ12-2に対しメッセージ②を送る。このメッセージ②はライト要求により不揮発性キャッシュメモリ48-2の更新が必要であり、ライトデータを共用バス18を経由してリモート側の不揮発性キャッシュメモリ48-2に転送することを通知する。

【0092】このメッセージ②を受けたりモート側のコントローラ12-2は自己のキャッシュ管理テーブルを更新し、不揮発性キャッシュメモリ48-2内に書込用の空間を割り当て、レディ応答をローカル側のコントローラ12-1に返す。リモート側コントローラ12-2からのレディ応答をステップS10で判別すると、ロー

カル側のコントローラはホストコンピュータ10からのライトデータ③を自己の不揮発性キャッシュメモリ48-1、即ちローカル側のキャッシュメモリ48-1に書き込み、同時にステップS2でリモート側のキャッシュメモリ48-2に共用バス18経由で転送して書き込ませる。

【0093】この結果、ライトデータはローカル側およびリモート側のキャッシュメモリ48-1、48-2の両方に書き込まれる。ステップS12でローカル側およびリモート側のキャッシュメモリ48-2に対するライトデータの転送書き込みが済むと、この時点でライトバックは行わず、ステップS13でライト終了応答をホストコンピュータ10に転送して、一連のライト要求に対する処理を終了する。

【0094】図6は信頼性重視モードにおけるライトバックの処理動作を示したフローチャートである。図5のライト要求に対する処理に示したように、ローカル側およびリモート側のキャッシュメモリにライトデータが書き込まれた時点ではディスク装置に対する書き込み所謂ライトスルーは行わず、予め定めたライトバック条件が成立したときにディスク装置に書き戻すライトバック処理を実行する。

【0095】このライトバック処理は、まずステップS1でコントローラ12-1がアイドル状態か否かチェックしており、アイドル状態になればステップS6で要求された事項の処理を行っているが、アイドル状態に入るとステップS2でライトバック条件の成立の有無をチェックする。ここでライトバック条件とはキャッシュ管理テーブルに設けているLRUテーブルに基づきキャッシュデータの追い出し条件がスケジュールされている場合である。

【0096】即ち、キャッシュ管理テーブルのLRUテーブルには、最新に使用されたキャッシュブロックを先頭とし、最も使用されていない時間のたったキャッシュブロックを末尾とするキャッシュブロックの登録が行われており、例えば一定時間使われなかったテーブル末尾のキャッシュブロックがキャッシュ無効化されるが、この内ディスク装置への書き込みが済んでいないデータ、所謂ダーティデータについてはライトバックがスケジュールされる。

【0097】ステップS2でライトバック条件の成立を判定すると、ライトバック対象となっているLRUテーブルの末尾から一定範囲のキャッシュブロックを読み出してディスク装置30に書き込むライトバックを実行する。例えば図8に示したように、先行するライト処理でコントローラ12-1がホストコンピュータ10からライト要求①を受けて自己の不揮発性キャッシュメモリ48-1にライトデータを書き込んでおり、その後のアイドル状態でライトバック条件が成立した場合には、④に示すようにディスク装置30にデータを書き込むライト

バック処理を行う。

【0098】再び図6を参照するに、ステップS3でディスク装置30に対するライトバックが済むと、ステップS4でライトバックが済んだキャッシュブロックをキャッシュ管理テーブルのハッシュテーブルから削除するキャッシュ無効化を行う。更に、ローカル側のコントローラ12-1はステップS5で共用バス18を経由してリモート側のコントローラ12-2にメッセージを送り、ライトバックが済んだデータのキャッシュ管理テーブルからの削除を指示し、リモート側についてもキャッシュ無効化を行う。

4. 容量重視モードのキャッシュ制御

この性能重視モードは2台のコントローラ12-1, 12-2に設けた各キャッシュメモリ48-1, 48-2は不揮発性とし、各キャッシュメモリ498-1, 48-2のデータを独立とする。そのため、2つのキャッシュメモリ48-1, 48-2に同一データを格納した信頼性重視モードに比べ、最大で2倍のキャッシュ容量とすることができる。換言すると、性能重視モードは容量重視モードと言うことができる。

【0099】リード処理はリード要求を受けたローカル側でミスヒットとなった場合にはリモート側にキャッシュ処理を依頼し、リモート側でキャッシュヒットになれば、直接、要求データをホストコンピュータにデータ転送する。ディスク装置からのステージングはローカル側およびリモート側の両方でミスヒットとなった場合にディスク装置からローカル側に対してのみ行う。またライト処理については、ライト要求を受けたローカル側のキャッシュメモリに対してのみ行う。

【0100】更にライトバックについても、ライトバック条件が成立したローカル側のキャッシュメモリからディスク装置に書き込むライトバックのみを行い、信頼性重視モードにおけるライトバック終了後のリモート側に対するキャッシュ無効化の指示は行わない。図9は性能重視モードにおけるキャッシュ動作を示したフローチャートである。ステップS1でホストコンピュータ10からアクセス要求があると、ステップS2でアクセスモードを判別し、リード要求であればステップS3で自己のキャッシュメモリ48-1、即ちローカルキャッシュメモリ48-1のヒットの有無を判定する。

【0101】ローカルキャッシュメモリ48-1のヒットが判別されると、ステップS4で要求データをホストコンピュータ10にデータ転送する。ローカル側キャッシュメモリ48-1でミスヒットとなった場合には、ステップS5で共用バス18を経由してリモート側のコントローラ12-2にキャッシュ処理を依頼する。このローカル側からのキャッシュ処理の依頼を受けたリモート側のコントローラ12-2は、リモート側キャッシュメモリ48-2でキャッシュヒットになれば要求データを直接ホストコンピュータ10にデータ転送し、応答結果

をローカル側のコントローラ12-1に通知してくる。

【0102】一方、リモート側でもキャッシュミスヒットとなった場合には、ステップS6でリモート側からのヒット応答が得られないことからステップS7に進み、ディスク装置30からローカル側のキャッシュメモリ48-1に要求データを書き込むステージングを行う。ステージングが済むと、ステップS8でホストコンピュータ10へのデータ転送を行い、これ以降ステージングされたデータはヒット可能状態となる。このような一連の処理を終了すると、リモート側のコントローラ12-1はステップS9でホストコンピュータ10にリード終了応答を返す。

【0103】図11は図9のステップS5～S8に示したローカル側およびリモート側の両方の不揮発性キャッシュメモリ48-1, 48-2がホストコンピュータ10からのリード要求①に対しミスヒットとなったときの動作を示している。即ち、ホストコンピュータ10からのリード要求①を受けたローカル側のコントローラ12-1は自己の不揮発性キャッシュメモリ48-1のミスヒット②を判定すると、共用バス18を経由してリモート側のコントローラ12-2にメッセージを送り、リモート側の不揮発性キャッシュメモリ48-2に要求データが存在するか否かの調査を依頼する。

【0104】リモート側のコントローラ12-2は自己のキャッシュ管理テーブルのハッシュテーブルを参照し、もし要求データが存在すればヒット応答と要求データが存在するアドレスをローカル側のコントローラ12-1に通知する。これに基づき、ローカル側のコントローラ12-1はリモート側の不揮発性キャッシュメモリ48-2から要求データ⑤を読み出してホストコンピュータ10に転送する。

【0105】尚、図11にあっては、リモート側の不揮発性キャッシュメモリ48-2から読み出した要求データ⑤をコントローラ12-2から直接ホストコンピュータ10にデータ転送しているが、共用バス18からローカル側のコントローラ12-1を経由してホストコンピュータ10にデータ転送をすることもできる。しかし、リモート側のコントローラ12-2から直接転送した方が、新たなホストコンピュータ10とコントローラ12-2との間のチャネル結合を考慮しても高速に処理できる。

【0106】再び図9を参照するに、ステップS2でライト要求を判別した場合にはステップS10に進み、ローカル側のキャッシュメモリ48-1に対してのみライトデータを転送して書き込み、ステップS11でライト終了応答を返す。例えば図12に示すように、ホストコンピュータ10からコントローラ12-1に対しライト要求が行われたときには、自己のキャッシュ管理テーブルを更新して格納領域を確保し、ライトデータ②を不揮発性キャッシュメモリ48-1に転送して書き込む。こ

の段階ではディスク装置30に対する書込みは行わず、その後のライトバック処理で書き込む。

【0107】図10のフローチャートは性能重視モードにおけるライトバック処理を示す。このライトバック処理にあっては、ステップS1でアイドル状態を判別すると、ステップS2でライトバック条件の成立の有無をチェックし、ライトバックがスケジュールされた条件成立であるとステップS5に進み、図12の③に示すように、ローカル側の不揮発性キャッシュメモリ48-1からライトバック対象データを読み出してディスク装置30に書き込み、ステップS4でライトバックが済んだキャッシュブロックをキャッシュ管理テーブルから削除するキャッシュ無効化を行う。

5. 平均モードのキャッシュ制御

この平均モードで使用する各系統のキャッシュメモリは、共に不揮発性キャッシュメモリとする必要がある。平均モードは信頼性重視モードと性能重視モードの混合型となる。

【0108】概略的には、リード要求に対するミスヒット時のステージングはローカル側のキャッシュメモリに対してのみ行い、これは性能重視型と同じである。またライト要求に対するライトデータの書込みはローカル側およびリモート側の両方のキャッシュメモリに対して行い、キャッシュメモリに同一データを格納した二重化を図る。これは性能重視モードと同じである。

【0109】このような信頼性重視モードと性能重視モードの混合型である平均モードのキャッシュ制御にあっては、ライト要求が少なくリード要求の多い環境では、キャッシュメモリの容量は性能重視モードと同様に2台のキャッシュメモリを設けたことで2倍となる。またライト要求が多く発行されリード要求の少ない環境では、信頼性重視モードと同様に片側データの送出時でもデータを保障することができる。

【0110】図13のフローチャートは平均モードにおける処理動作を示している。まずステップS1でホストコンピュータ10からのアクセス要求があるとステップS2でアクセスモードを判別し、リード要求であればステップS3で自己のキャッシュメモリ48-1、即ちローカル側のキャッシュメモリ48-1に要求データが存在するか否かのヒット判定を行う。

【0111】キャッシュヒットが得られると、ステップS4で要求データを読み出してホストコンピュータ10に転送する。ローカル側のキャッシュメモリ48-1がミスヒットとなった場合には、ステップS5でリモート側のコントローラ12-2へ共用バス18を経由してキャッシュ処理を依頼する。このキャッシュ処理の依頼を受けたリモート側のコントローラ12-2より、リモート側のキャッシュメモリ48-2でのキャッシュヒットおよび要求データのアドレスを示す応答を受けると、リモート側のキャッシュメモリ48-2より要求データを

読み出して、ステップS8でホストへのデータ転送を行う。

【0112】一方、ステップS6でリモート側よりミスヒットの応答を受けた場合には、ステップS7でディスク装置30からローカル側のキャッシュメモリ48-1に対するステージングを行った後、ステップS8でホストコンピュータ10へデータを転送する。これら一連のリード処理を終了すると、ステップS9でリード終了応答を返す。

【0113】一方、ステップS2でアクセス要求を受けたローカル側のコントローラ12-2がライト要求を判別した場合には、ステップS10でリモート側のコントローラ12-2に共用バスを経由してメッセージを送ってライトデータの転送を通知する。この通知を受けて、リモート側のコントローラ12-2はキャッシュ管理テーブルを更新してライトデータを書き込むための領域を確保し、ローカル側にレディ応答を返す。

【0114】ステップS10でリモート側よりレディ応答があると、ステップS12でローカル側のコントローラ12-1は自己のキャッシュメモリ48-1、即ちローカル側のキャッシュメモリ48-1にライトデータを転送して書き込み、またステップS3でリモート側のキャッシュメモリに共用バスを経由してライトデータを転送して書き込ませる。

【0115】最終的にローカル側およびリモート側のキャッシュメモリ48-1、48-2に対する書込みが終了すると、ステップS12でライト終了応答を返す。このようなライト処理によりローカル側およびリモート側のキャッシュメモリ48-1、48-2に同一データが書き込まれることになる。図14は平均モードにおけるライトバック処理を示したもので、図6に示した信頼性重視モードのライトバック処理と同じである。

【0116】即ち、ライトバック対象となったローカル側のキャッシュメモリ48-1のダークデータデータをディスク装置30に書き込むライトバック処理を行い、ライトバックが済んでからローカル側のキャッシュメモリ48-1からライトバック済みデータを削除するキャッシュ無効化を行い、更に共用バス18でリモート側にメッセージを送って、リモート側のキャッシュメモリ48-2ライトバック済みのデータを削除するキャッシュ無効化を行わせる。

6. 第2発明の全体機能

図15は図2に示したディスクアレイを対象とした第2発明によるキャッシュ制御の機能を示し、図2のコントローラ12-1側を代表して示している。

【0117】コントローラ12-1にはキャッシュ制御部46、不揮発性キャッシュメモリを用いたキャッシュメモリ48、キャッシュ管理テーブル50およびディスクアレイ制御部52が設けられている。キャッシュ制御部46にはリードキャッシュ制御部46-1、ライトキ

キャッシュ制御部46-2およびライトバック制御部45の各機能が設けられる。

【0118】キャッシュ管理テーブル50はハッシュテーブル60、LRUテーブル62および空スペース管理テーブル64で構成される。ディスクアレイ制御部52には、この実施例にあってはRAID0、RAID1、RAID3およびRAIDの形態をもつ動作機能が予め設けられている。RAIDの動作形態の指定はホストコンピュータ10からの指示で論理的に行うことができる。

【0119】ディスクアレイ28は図2に示したハードウェア構成に対応するものであり、縦方向の並びをランクと読んでランク番号R0～R3で示し、横方向の並びをポートと読んでポート番号P0～P5で示している。図16は図15のキャッシュメモリ48の構造を示す。キャッシュメモリ48はデータアクセスの最小単位であるキャッシュブロック54から構成されている。キャッシュブロック54のサイズは、右側に取り出して示すようにホストコンピュータ10がアクセスする最小単位であるホストブロック（論理ブロック）の整数倍となる。

【0120】図16にあっては、ホストブロック54の4倍にキャッシュブロック54を設定した場合を示している。このキャッシュブロック54のサイズ設定は通常、システム側で行われ、ユーザは介在しない。図17はディスク装置30のデータブロックに対するホストブロックとの関係を示している。ディスクアレイシステムにおけるディスク装置30のフォーマットは固定長ブロック形式（FBA）を採用しており、デコード長が可変のカウントキーデータ形式（CKD）はデータのストライピングやパリティの生成が複雑になることから採用しない。

【0121】磁気ディスク30の中に示す両側のインデックスで挟まれた各セクタを構成する固定長データブロック58は、通常、512バイトとなる。ホストブロック56はディスク装置30の固定長データブロック58の整数倍となり、この例ではホストブロック56はデータブロック58の4倍であることから2,048バイトとなる。

【0122】図18は図15のキャッシュ管理テーブル50に設けたハッシュテーブル60の構造を示す。ハッシュテーブル60はハッシュエントリ66とハッシュエントリで指定される登録データ68で構成される。ハッシュエントリ66はホストコンピュータ10からのアクセス要求で得られたデバイス情報をハッシュパラメータとして計算したエントリ値（ハッシュアドレス）で指定され、指定されたハッシュエントリにキャッシュメモリ上に存在するキャッシュブロックCBの登録状態が登録データ68として図示のように登録されている。

【0123】このハッシュテーブル60はホストコンピュータから要求のあったホストブロックがキャッシュメ

モリ上に存在しているかどうかの判定、即ちヒット／ミスヒット判定に主に使用される。図19は図15のキャッシュ管理テーブル50に設けたLRUテーブル62を示している。LRUテーブル62はキャッシュメモリ48上に格納されたキャッシュブロックをアクセス順にリンクしている。

【0124】例えば新たなキャッシュブロックの書き込みが行われると、LRUテーブル62の先頭に最新に使用されたキャッシュブロックとしてリンクする。またキャッシュブロック上に存在する任意のキャッシュブロックがリード要求に対し読み出された場合にも、最新に使用されたキャッシュブロックとしてLRUテーブル62の先頭にリンクされる。

【0125】このようなLRUテーブル62はキャッシュメモリ48上に存在するデータの追出しと、ディスク装置にキャッシュメモリ上のデータを書き込むライトバック制御に使用される。即ち、LRUテーブル62の最後にキャッシュメモリ48上に存在している時間が最も長い、長時間アクセスされていないキャッシュブロックがリンクされていることから、この最後にリンクされたキャッシュブロックが次に追い出されるキャッシュブロックの候補となる。

【0126】また追出し候補となったキャッシュブロックがディスク装置への書き込みが済んでいないデータ所謂ダーティデータであった場合には、ディスク装置に書き戻すライトバック処理がスケジュールされ、ライトバック後に追出しを行うことになる。

7. キャッシュ空スペースの確保

図20はキャッシュメモリ48におけるキャッシュブロックの使用状態の一例を示している。

【0127】キャッシュメモリ48において、キャッシュブロックRはライトバックによりディスク装置に同一データが格納されているブロックを示す。キャッシュブロックWはキャッシュメモリ48上で書込処理は行われたがライトバックが済んでおらず、ディスク装置に同一データが存在しないダーティデータが存在するブロックを示す。また「空」は未使用のキャッシュブロックである。

【0128】ディスクキャッシュ制御にあっては、ホストコンピュータ10からキャッシュメモリ48上に存在しないデータに対してのリード要求またはライト要求が指示された場合、未使用の空間から新たなキャッシュブロックを割り当て、リード要求に伴ってディスク装置30からステージングされたデータあるいはホストコンピュータから受領したライトデータを書き込む。

【0129】図20のキャッシュメモリ48における未使用の空キャッシュブロックの空間は、図21に示す空スペース管理テーブル64により管理される。空スペース管理テーブル64は空キャッシュブロックの数をエントリ番号1, 2, 3, ……Nとし、各エントリ番号ご

とに連続する空ブロックの先頭のブロック情報を登録している。なお、テーブル内容は説明の都合上示したもので、実際には設けられない。

【0130】ここで図20のキャッシュブロック48におけるX方向のアドレスを $X=00\sim07$ 、Y方向のアドレスを $Y=00\sim n$ で表すと、空スペース管理テーブル64における先頭ブロックのアドレスは「CBXY」で示される。例えば空スペース管理テーブル64の空ブロック数1個を示すエントリ1には、図20のキャッシュメモリ48におけるブロックアドレスCB0103、CB0004の2つの空ブロックが登録されている。空ブロック2個分を示すエントリ番号2については、2個空ブロックが連続した先頭ブロックのアドレスCB0102が登録されている。

【0131】このような空スペース管理テーブル64を設けることで、新たなキャッシュブロックの割当要求に対してはなるべく物理的に連続したキャッシュブロックを割り当てている。更に空スペース管理テーブル64で管理されるキャッシュメモリ48の空スペースは、常に一定量が残るように管理される。

【0132】即ち、キャッシュメモリ48の空スペースの残り容量が規定値以下になるような状況が発生すると、図19に示したLRUテーブル62を参照してキャッシュメモリ上の有効データを追い出すためのスケジュールが行われる。この追出し対象となったキャッシュブロックにつき、ディスク装置にも同一データが存在するキャッシュブロックRの部分については、図18に示したバッシュテーブル60からの登録を抹消するキャッシュ無効化を行う。一方、ディスク装置にデータが存在しないダーティデータのキャッシュブロックWの部分については、ディスク装置へのライトバックがスケジュールされる。

【0133】このようなキャッシュメモリ48の残り容量を規定値以上に保つ処理により、常にキャッシュメモリ上に規定の空き空間が準備されており、従って新たなデータ書き込みを伴うリード要求およびライト要求に対し、直ちに処理を開始することができる。

8. RAID動作形態の制御

図15に示したディスクアレイ制御部52はホストコンピュータ10からの論理デバイスの指定に基づき、RAID0、1、3または5のいずれかの形態のアクセス動作をディスクアレイ28に対し実行する。

【0134】図22はランク数4、ポート数6のディスクアレイ28に対するホストコンピュータが指定する論理デバイス番号と、RAID動作のためのディスク装置の形態を示している。まずランクR0に設けられた6台のディスク装置の内、ポートP0～P4の5台がホストコンピュータ10における論理デバイス番号0で指定され、RAID3の動作形態をとる。この内、ポートP0～P3のディスク装置はデータ用であり、ポートP4の

ディスク装置がパリティ用に固定的に決められる。

【0135】尚、ランクR0のポートP5のディスク装置は予備機として設けられたホットスベアHSに割り当てられる。次のランクR1のポートP0～P5の5台のディスク装置はホストコンピュータ10における論理デバイス番号1で指定され、RAID5の動作形態をとる。RAID5の動作形態にあつては、ポートP0～P4の中の4台のディスク装置がデータ用、1台がパリティ用となる。パリティ用のディスク装置はセクタ位置が替わるごとに循環する。

【0136】ランクR2のポートP0、P1およびポートP2、P3に設けられた2台を1組としたディスク装置は、ホストコンピュータ10における論理デバイス番号2、3で各々指定され、RAID1の動作形態、即ち2台のディスク装置に同一データを格納したミラーディスクとしての動作形態をもつ。更にランクR3のポートP0～P3に設けた4台のディスク装置はホストコンピュータ10における論理デバイス番号4で指定され、RAID0の動作形態をとる。即ち、データをストライピングして並列的に書き込む動作は論理デバイス番号0のRAID3の動作形態と同じであるが、冗長性を確保するためのパリティディスクを備えていない。

【0137】図23はRAID3の動作形態におけるキャッシュブロック、ストライピングおよびディスク格納状態を示している。まずキャッシュブロック54は、この例では4つのホストブロックL0～L3で構成されている。RAID3の動作モードでディスク装置30-00～30-04に書き込むために、例えばホストブロックL0について取り出して示すように、所定数の固定長データブロックのサイズで16分割したバイトデータL0-1～L0-16にストライピングする。

【0138】このようなストライピングデータについて、データ用のディスク装置30-00～30-03の4台数4台に対応したストライピングデータ単位にパリティを計算する。そして4つのストライピングデータに計算したパリティを加えた5つのバイトデータを、5台の磁気ディスク装置30-00～30-04に並列的に書き込む。

【0139】図24はRAID5の動作形態におけるキャッシュブロック、データストライピングおよびディスクアレイ格納状態を示している。この例では説明を簡単にするため、ディスクアレイ28に並列書き込まれるホストブロック数にキャッシュブロックのサイズを一致させている。キャッシュメモリ48上には並列書き込み可能なデバイス、即ちディスク装置30-00～30-03の4台に対応したホストブロック数4のサイズをもつキャッシュブロック54-1～54-4が設けられ、図示のように連続するキャッシュブロック54-1～54-4に、同じく連続するホストブロックL0～L15が格納されていたとする。

【0140】このような場合には、例えばキャッシュブロック54-1を取り出し、キャッシュブロック54-1を構成する4つのホストブロックL0~L3からパリティを計算し、ホストブロックL0~L3にパリティP1を加えた5つのデータを並列的にディスク装置30-00~30-04に書き込む。通常、RAID5の動作形態にあつては、データ長の短いホストブロック1個単位のアクセスが行われる。この通常のRAID5の動作形態における書き込みは、例えば図示のようにディスクアレイ28側にホストブロックL0~L3およびパリティP1が格納されている状態で、新たに論理ブロックL0のみを書き込む場合には、次のようなリードモディファイライトを実行する。

【0141】まずディスク装置30-00および30-04により旧ブロックL0と旧パリティP1を読み出す。これを(L0)old, (P1)oldとする。続いて中間パリティ(P1)intを

$$(P1)int = (P1)old + (L0)old$$

として、排他論理和を求める。続いて中間パリティ(P1)intと新ブロック(L0)newから新パリティ(P1)newを

$$(P1)new = (P1)int + (L1)new$$

として求める。新ブロック(L0)newと計算で求めた新パリティ(P1)newをディスク装置30-00, 30-04に対し並列的に書き込む。

【0142】このようなRAID5の動作形態における並列書き込み可能なディスク台数以下のホストブロックの書き込みについては、旧パリティと旧データの読出しを伴うことからディスク2.5回転分の処理時間がかかり、これがライトペナルティとなる。第2発明にあつては、RAID5の動作形態にあつても可能な限りRAID3の動作形態によるディスクアレイに対するライトアクセスを行って、ライトペナルティの低減を図る。

【0143】図25はRAID1の動作形態におけるキャッシュブロックとディスク格納状態を示している。RAID1は2台のディスク装置30-20, 30-21に同一データを書き込むミラーディスクとしての使い方であり、キャッシュブロック54からホストブロックL0~L3ごとに順次取り出してディスク装置30-20, 30-21のそれぞれに並列的に書き込む。

【0144】図26はパリティディスクをもたないRAID0の動作形態のキャッシュブロック、ストライピングおよびディスク格納状態を示す。尚、図24のRAID5の場合と同様、説明を簡単にするため、4台の磁気ディスク装置30-30~30-34に並列書き込みする論理ブロック数4に一致するキャッシュブロックのサイズとした場合を例にとっている。

【0145】このRAID0の動作形態では、キャッシュブロック54を構成する4つのホストブロックL0~L3を、4台のディスク装置30-30~30-34の

それぞれに割り当てて並列的に書き込み、この場合にパリティディスクはもたないことから、パリティの計算は行わない。

9. ステージング処理の高速化

ディスクキャッシュ制御にあつては、ホストコンピュータ10からのリード要求に対しキャッシュメモリ48上に要求データが存在せずにミスヒットとなった場合には、ディスク装置から要求データを読み出してキャッシュメモリ上に書き込むステージングを行う。

【0146】本発明にあつては、RAID0, 1, 3および5のいずれの動作形態にあつても、ディスクアレイ28の同時アクセス可能な全ディスク装置からの並列リードによるステージングで要求データのキャッシュブロックに対するステージングを高速で行う。図28はRAID5の動作形態を例にとった本発明のステージングを示している。いま16ホストブロックのサイズをもつキャッシュブロック54-1がキャッシュメモリ48上に展開されており、このキャッシュブロック54-1のうちの最後の2つのホストブロックL14, L15のみが有効データであり、またホストブロックL14, L15はディスク装置への書き込みが済んでいないダークデータであつたとする。

【0147】ここで、ホストコンピュータ10より論理ブロックL4~L7を指定したリード要求①がホストコンピュータから行われたとする。このリード要求に対し、キャッシュメモリ上に要求データが存在しないことからミスヒットとなり、ステージングを行う。ステージングはキャッシュブロック54-2に示すように、ディスクアレイ28のディスク装置30-10~30-12, 30-14の4台をアクセスしてホストブロックL4~L7をリードしてキャッシュブロック54-2上に書き込み、ステージングが完了するとヒット状態となることから、要求データをホストコンピュータ10に転送する。

【0148】このようなステージングが終了した後に本発明にあつては、それ以降のアクセス要求に対するキャッシュメモリのヒット効率を向上してキャッシュ効果を最大限に発揮するため、プリフェッチ動作を行う。即ち、キャッシュブロック54-3に示すように、ステージングが済んだホストブロックL4~L7に後続する並列アクセス可能なディスク数4に一致する4つのホストブロックL8~L11をディスク装置30-10, 30-11, 30-13, 30-14から読み出してプリフェッチする。

【0149】RAID5にあつては、基本的にはリード転送速度はデバイス速度以上は出ることがなく、またステージング速度がRAID3より遅く、少量データのリード処理が中心になることから、プリフェッチ量は次のパリティグループまでとする。ここで、パリティグループとはパリティディスクを同一とする各ディスク装置の

データブロックを意味する。例えば、図27(E)におけるデータブロックL0~L3, L4~L7などのことを意味する。

【0150】このようにRAID5にあつては、プリフェッチ量は次のパリティグループまでとなるが、キャッシュブロックサイズをパリティグループのデータブロックサイズと同一になるようにストライピングを行えば、即ち図24に示したストライピングを行えばキャッシュブロックの中にステージングしたデータブロックに続くキャッシュブロックの最後までデータブロックがプリフェッチ量となり、このプリフェッチをステージングに含めてRAID3的なステージングが可能となる。

【0151】一方、RAIDでは図23に示したように、1つのホストブロックが各ディスク装置に分散して格納されており、ステージングが高速にできること、および大容量のデータが要求される場合が多いことを考慮し、アクセス要求が行われたステージングデータを含むキャッシュブロックの最後までデータブロックをプリフェッチする。

【0152】再び図27を参照するに、キャッシュブロック54-3に召すステージングしたホストブロックL4~L7に続くブロックL8~L11をプリフェッチした状態にあつても、キャッシュブロック上の有効データは2箇所に分散している。このようなステージングに伴うプリフェッチ後の状態でキャッシュブロック上の有効データが分散している場合には、キャッシュブロック54-4に示すようにホストブロックL12, L13に対応するデータブロックをディスクアレイ28のディスク装置30-12, 30-13から読み出して内部的にステージングし、キャッシュブロック上で有効データを連続データとする。

【0153】このようにキャッシュブロック上の有効データを連続データとすることで、キャッシュブロック上での以後のデータ管理を容易にする。例えば、キャッシュブロック上で有効データが複数箇所に分散して存在していたとすると、最終的には分散した各ホストブロックに対応したマップ管理が要求される。これに対し、キャッシュブロック上に連続した領域を有効データとすることで、有効データの一部が書き替えられた際のライトバック処理において、RAID3的な並列書き込みによる高速化を図ることができる。

【0154】10. 大量シーケンシャルデータの追出し促進

一般にキャッシュ制御にあつては、キャッシュメモリ48上の同一データを頻繁にリードするとキャッシュ効果を発揮することができる。これに対し、ホストコンピュータ10がキャッシュメモリ48上に存在しない大量データのリード要求を行ったとき、ミスヒットに伴ってディスクアレイ28から要求された大量データをキャッシュメモリ48上にステージングしてホストコンピュータ

10に転送するようになる。

【0155】このようにステージングされた大量のデータがその後、参照されないにも関わらず、そのままキャッシュメモリ48上に残されていると、キャッシュスペースが有効に利用されず、ヒット効率が低下することになる。しかしながらLRUテーブル62を用いたキャッシュメモリ48の管理では、一度しか参照されなかった大量データであっても直ちに無効化することはできない。

【0156】そこで本発明のディスクキャッシュ装置にあつては、ホストコンピュータ10からの一定量を越えるデータのリード要求に伴って大量データのステージングが行われた場合には、本来、LRUテーブル62の先頭にステージングされたキャッシュブロックを登録すべきところ、LRUテーブル62の終端側に近い位置にステージングされた大量データのキャッシュブロックを登録する。

【0157】図29のフローチャートは大量データのステージング処理を示す。まずステップS1でホストコンピュータ10からのリード要求に伴うミスヒットでディスクアレイ28側からのステージングが行われたならば、ステップS2で、ステージングされたデータ量が所定値例えば100キャッシュブロック以上か否かチェックする。

【0158】100キャッシュブロック未満であれば通常通り、ステップS3でLRUテーブル62の先頭に登録する。一方、100キャッシュブロック以上であった場合には、ステップS4でLRUテーブル62の終端側の近くに登録する。このようなステージングされた大量データについて、LRUテーブル62の終端側に強制的に登録することで、一度しか参照されないステージングされた大量データは比較的短い時間の経過後にハッシュテーブル60から削除されて追い出されることになり、キャッシュスペースが不要に占有されてしまうことを防止できる。

11. パリティデータのステージング

本発明のディスクキャッシュ装置におけるRAID5の動作状態において、原則的にはキャッシュメモリ48上にパリティデータのステージングは行っていない。しかしながらRAID5の動作状態にあつては、ディスクアレイに対する書き込みに新パリティを計算するためにディスクアレイ28から旧データおよび旧パリティを読み出す処理を必要とするため、処理時間が長くなるライトベナリティをもつ。

【0159】そこで本発明のステージング処理の他の実施例にあつては、オプション指定によりパリティデータのステージングを可能とする。図30のフローチャートはパリティデータのステージングを可能とする本発明のステージング処理の他の実施例を示す。このステージング処理にあつては、ホストコンピュータ10からのリー

ド要求に対するキャッシュメモリ48のミスヒットでステージング処理が起動されると、まずステップS1で動作形態がRAIDであり且つパリティデータのステージングがオプション指定されているか否かチェックする。

【0160】パリティデータのステージングを行わせるオプション指定は、ホストコンピュータ10からのコマンドあるいはメンテナンス端末からの指示によりセットされる。ステップS1でパリティデータのステージングを示すオプション指定が有効であるとステップS2に進み、パリティデータを含むデータをディスクアレイ28からキャッシュメモリ48上にステージングする。ステップS1でオプション指定がなければステップS3に進み、通常のパリティを含まないデータのディスクアレイ28からのステージングを行う。

【0161】図31はパリティデータをステージングしない場合とステージングした場合のライトバック処理を対比して示す。図31(A)はパリティデータをステージングしていない場合であり、キャッシュブロック54-10のホストブロックL1のみが有効データであったとする。

【0162】このキャッシュブロック54-10の有効データであるホストブロックL1をディスクアレイ28にライトバックする場合には、図24のRAID5におけるディスク書込みに示したようにディスク装置30-11から対応する旧データを読み出し、またディスク装置30-14から旧パリティを読み出し、中間パリティを生成した後新データから新パリティを作成し、最終的にディスク装置30-11、30-14に新データと新パリティを書き込む処理を必要とする。

【0163】このため、キャッシュブロック54-10上の有効ブロックL1をディスクアレイ28に書き込むためにはディスク装置の2.5回転分の時間が必要となる。これに対し図31(B)に示すキャッシュブロック54-10上にパリティデータP1をステージングしていた場合には、有効ブロックL1の更新に際しキャッシュブロック54-20上で新たなパリティデータP1を計算でき、ディスクアレイ28側をリードアクセスすることなく直接、更新ブロックL1と新パリティデータP1をディスク装置30-11、30-14に書き込むことができる。

【0164】このため、ライトバック処理はディスク装置1.5回転分の時間で済み、通常RAID5の動作状態のディスク書込み処理におけるライトペナルティは発生しない。勿論、パリティデータをキャッシュメモリ48にステージングした場合には、その分だけキャッシュメモリ48の有効利用領域が低減することになる。

12. ライトバックデータの並び替えによる最適化
図22の空スペース管理テーブル64で説明したように、キャッシュメモリ48上の空スペースが一定値以下になった場合は、図19に示すLRUテーブル62を参

照し、使用されずに最も長い期間格納されているキャッシュブロックを追い出すためのライトバック処理がスケジュールされる。

【0165】図32はライトバック処理がスケジュールされる際のキャッシュブロック62の一例を示している。通常キャッシュ制御にあつては、LRUテーブル62の最後にリンクされているキャッシュブロックCB-fをライトバック対象とするスケジュールを行う。これに対し本発明のキャッシュ制御にあつては、ライトバック条件が成立したとき、即ちキャッシュメモリ48の空ブロック数が規定の閾値以下となったときには、LRUテーブル62の最後のキャッシュブロックから所定ブロック数 α だけ遡った範囲をライトバック対象範囲68として指定するスケジュールを行う。

【0166】この場合にはキャッシュブロックCB-a~CB-fの6つがライトバック対象範囲68に含まれるように指定される。ここでライトバック対象範囲68に含まれる斜線で示すキャッシュブロックCB-b、CB-e、CB-fの3つが、図33に示すホストブロックを含んでいたとする。即ち、キャッシュブロックCB-bはホストブロックL4~L7の対応データを含み、キャッシュブロックCB-eはホストブロックL8~L11の対応データを含み、キャッシュブロックCB-fはホストブロックL0~L3の対応データを含んでいたとする。

【0167】このようなホストブロック対応データを見ると、キャッシュブロックCB-f、CB-b、CB-eの順番に並び替えることで、ホストブロックL0~L11の対応データを連続データとすることができる。ホストブロック対応データを連続データとするライトバック対象範囲68に含まれるキャッシュブロックの並び替えを行った場合には、ディスクアレイ28に各キャッシュブロックをライトバックする際にRAID3的な並列書込みを高速で実行できる。

【0168】即ち、キャッシュブロックCB-f、CB-b、CB-eに含まれる4つの論理ブロック対応データのそれぞれからパリティP1、P2、P3を計算し、順番に4つのデータとパリティを並列的にディスクアレイ28のディスク装置30-11~30-14に連続的に書き込む高速ライトバック処理が達成できる。このようなライトバック対象範囲68に含まれるキャッシュブロックの論理ブロック対応データを連続するような並び替えによるライトバックが完了すると、図32に示したLRUテーブル62は図34の状態となる。

【0169】図35のフローチャートはキャッシュブロックの並び替えを伴うライトバックのスケジュールが行われるライト処理を示し、図28のステップS13に示したライト処理の詳細を示すことになる。まずステップS1でホストコンピュータ10からのライト要求に伴って、ライトキャッシュ制御部46-2でキャッシュメモ

り48上に確保された8ブロックに対しライトデータを書き込む。

【0170】ライトデータの書き込みが済むと、ステップS2で図22に示した空スペース管理テーブル64を参照してキャッシュメモリ48の空ブロック数を求め、空ブロック数が予め定めた閾値以下か否かをチェックする。閾値以下であればステップS3に進み、図32に示したようにLRUテーブル62の末尾から遡った所定数 α のブロックを含む範囲をライトバック対象範囲68として指定する。

【0171】ライトバック対象範囲68に含まれるキャッシュブロックがディスクアレイ28への書き込みが済んでいないライトバック対象となるダーティデータか、既にディスクアレイ28への書き込みが済んでいるライトバック対象から除外される非ダーティデータか否かをチェックする。もしライトバック対象範囲68に非ダーティデータのキャッシュブロックが含まれていれば、ステップS5で非ダーティデータのキャッシュブロックをハッシュテーブル60から削除してキャッシュ無効化を行う。

【0172】続いてステップS6でライトバック対象となるダーティデータのキャッシュブロックの有無をチェックし、ダーティデータのキャッシュブロックが存在すると、ステップS7でライトバックをスケジュールする。このライトバックのスケジュールにあつては、図32、図33に示したようにライトバック対象範囲68に含まれるキャッシュブロックについて、もし可能であればホストブロックに対応するデータが連続データとなるようにキャッシュブロックの並び替えを行った後にライトバック対象としてスケジュールする。

【0173】尚、ライトバック処理のスケジュールは図35に示すライト処理の完了時のみならず、図28に示したリード要求でキャッシュメモリがミスヒットとなったときのディスクアレイからのステージング終了時についても同様にスケジュールされることになる。

13. ライトバック処理の最適化

図36はRAID5の動作形態におけるライトバック処理の一例を示している。ここでキャッシュブロック54のサイズは16ホストブロックであり、ホストブロックL0～L15に対応するデータを格納することができ、この内、ホストブロックL2～L9に対応するデータが有効データとして書き込まれ、ライトバック対象データ75となっている。

【0174】一方、ディスクアレイ28側はディスク装置30-10～30-14のそれぞれは1ホストブロックのサイズに相当するセクタ単位にデータを読み書きしており、同時並列的には4つのホストブロックをアクセスすることができる。キャッシュブロック54のライトバック対象データ75に含まれるホストブロックL2～L9対応データについて、ディスクアレイ28のストライピング状態に対応し、先頭のホストブロックL2、L

3と末尾のホストブロックL8、L9の各対応データについてRAID5モードによる書き込みを行う。

【0175】即ち、対応する旧データおよびパリティデータをディスクアレイ28から読み出して、書き込みを行う新データを用いて新パリティを計算した後に並列的に書き込むリードモディファイライトを行う。これに対し、ディスクアレイ28の並列書き込み可能な4台のディスク装置に対応する連続する4つのホストブロックL4～L7の対応データについては、RAID3モードによる書き込みを行う。即ち、ホストブロックL4～L7の対応データからパリティP2を計算し、直接ディスクアレイ28に書き込む。

【0176】このようなRAID5モードにおけるライトバック処理について、通常のRAID5モードに書き込みに加えてRAID3モードによる高速書き込みを組み合わせることで、最適化されたライトバック処理を行うことができる。図37はRAID5の動作状態におけるライトバック処理において、ライトバック対象データ75の中の連続するホストブロック対応データが、ディスクアレイ28側の並列書き込みディスク台数より僅かに少なかった場合の最適化処理を示している。

【0177】いまキャッシュブロック54の中にライトバック対象データ75としてホストブロックL1～L10の対応データが格納されていたとする。中央のホストブロックL4～L7の4つのデータについてはRAID3的な並列アクセスができる。これに対し、左側のホストブロックL1～L3対応データおよび右側のホストブロックL8～L11対応データは並列アクセスディスク台数4台に対し1つホストブロック対応データが少ない。

【0178】このような場合のライトバックにあつては、キャッシュブロック54-2に示すようにホストブロックL0対応データをディスクアレイ28のディスク装置30-10から①に示すようにリードしてステージングし、またホストブロックL11の対応データについてもディスクアレイ28のディスク装置30-14により②に示すようにリードしてステージングし、並列アクセス可能なディスク台数の整数倍となるホストブロック数の連続データとする。

【0179】このようなホストブロックL0～L11の連続対応データがキャッシュブロック54-2の上に得られれば、ホストブロックL0～L3、ホストブロックL4～L7、ホストブロックL8～L11の4つずつにストライピングして、それぞれパリティP1、P2、P3を求めた後、①④⑤に示すように並列的にディスクアレイ28に書き込むことで、RAID3的な高速ライトバック処理を行うことができる。

【0180】図38のフローチャートは図36、図37に示したRAID5のライトバック処理を含む本発明のキャッシュ制御における全体的なライトバック処理を示

す。まずステップS1でコントローラ12がアイドル状態か否かチェックし、アイドル状態にないときにはステップS14で要求事項の処理、即ち通常の処理を行っている。

【0181】アイドル状態が得られるとステップS2に進み、ライトバック処理がスケジュール済み、即ちライトバック条件が成立していればステップS3に進み、RAIDモードをチェックする。RAID5モードであればステップS4に進み、図32、図33に示したようなライトバック対象範囲におけるキャッシュブロックの連続ブロックへの並び替えを行う。なお、この連続ブロックへの並び替えは図35に示したライト処理の段階におけるスケジューリングとして行っておいてもよい。

【0182】続いてライトバック対象となった最初のキャッシュブロックを参照し、ディスクアレイ28側の並列書き込み可能なディスク台数に対応した全台数分のホストブロック対応データが有効か否かチェックする。もし全デバイス分のホストブロック対応データが有効であればステップS6に進み、RAID3モードによる書き込み処理を実行する。即ち、全デバイス分の対応データからパリティを計算した後、データおよびパリティを並列的に同時書き込みする。

【0183】一方、全デバイス分のキャッシュブロック対応データが有効でなかった場合には、ステップS7で不足データは1ディスク分か否かチェックする。不足データが1ディスク分であった場合には、ステップS8で、図37に示したように不足ブロックに対応するデータをディスクアレイからステージングし、ステージング後にステップS6に進んでRAIDモードでの書き込み処理を行う。

【0184】ステップS7で不足分が2ディスク分以上であった場合にはステップS9に進み、RAID5モードでの書き込み処理を行う。即ち、旧データおよび旧パリティをリードして中間パリティを計算し、更に新データから新パリティを計算し、新データおよび新パリティを並列的にディスクアレイに対し同時書き込みする。ステップS6またはステップS9の書き込み処理が済むと、ステップS10でライトバック対象となっているキャッシュブロックの全ブロックの処理が終了したか否かチェックし、終了していなければステップS4に戻り、同様な処理を繰り返し、終了していればステップS1で再び待機状態となる。

【0185】一方、ステップS3でRAID3の動作モードが判別された場合にはステップS11に進み、図23に示したストライピングおよび書き込みとなるRAID3モードの書き込み処理を行う。またステップS3でRAID0モードが判別された場合には、ステップS13で、図25に示すRAID0モードの書き込みを行う。更にステップS3でRAID1モードが判別された場合にはステップS13に進み、図26に示したパリティディスク

をもたないRAID1モードの書き込みを行う。

14. 統計的処理によるキャッシュブロックサイズの自動設定

本発明を含めたディスクキャッシュ装置を備えた計算機システムにおいて、ユーザが必ず設定を必要とするパラメータはホストブロックサイズ即ち論理ブロックサイズである。

【0186】これに対し、キャッシュブロックサイズの設定は通常、システム側で行っている。図16に示したように、キャッシュブロック54のサイズはホストブロック56が複数個集まったブロックとして構成される。このような従来の固定的なキャッシュブロックサイズのシステム側での設定に対し、本発明のディスクキャッシュ装置にあっては、最初のシステム立上げ時はシステム側が固定的にキャッシュブロックサイズを設定するが、次回以降の立上げ時にあってはユーザの使用環境に応じて自動的に最適なキャッシュブロックサイズに変更される。

【0187】図39のフローチャートは最適化されたキャッシュブロックサイズを自動設定するための統計処理を示している。最初のシステム立上げ時にあっては、システム側に固定的に設定したキャッシュブロックサイズによるキャッシュ制御が行われており、この状態でホストアクセスをステップS1で判別すると、ステップS2でリード要求またはライト要求が行われたホストブロックサイズを記録する。

【0188】ホストブロックサイズの記録処理はステップS5で規定周期に達するまで繰り返し行われている。ステップS3で規定周期例えば1カ月程度の時間に達するとステップS4に進み、ステップS2で記録されたホストブロックサイズから平均ブロックサイズを計算する。続いてステップS5で最適ブロックサイズの設定値を、ステップS4で求めた平均ブロックサイズに更新する。この最適キャッシュブロックサイズの更新が済むと、ステップS2で記録されたホストブロックサイズはメモリからクリアされ、次の周期の記録に入る。

【0189】この結果、システムの運用中に統計的処理で得られたキャッシュブロックサイズが最適ブロックサイズとして保持されることになる。そして一旦、システムの運用を止め、次に電源投入によるシステム立上げを行うと、図39の統計処理で求められている最適キャッシュブロックサイズに自動的にキャッシュブロックサイズを変更してキャッシュ制御を動作する。

【0190】図40はホストブロックサイズに対し最適なキャッシュブロックサイズの一例を示している。いま並列アクセスされるディスクアレイ28はディスク装置30-10~30-13の4台であり、それぞれのレコードにおける1セクタを構成するデータブロック58は512バイトであったとする。

【0191】この512バイトのデータブロック58に

対し、ホストブロックサイズが2倍の1,024バイトであったとすると、例えばホストブロックL0は対応するディスク装置30-10の2セクタ分に跨がる。キャッシュブロック54は4台の並列アクセス可能なディスク装置30-10~30-13のブロックデータサイズに一致するサイズをもつことがRAID3的なアクセスを可能とするために最適である。

【0192】従って、キャッシュブロック54は4つのホストブロックL0~L3に対応し、1つのホストブロックは1,024バイトとなることから、キャッシュブロックサイズは4,096バイトとすればよい。尚、上記の実施例は6ポート、4ランクのディスクアレイを例にとるものであったが、ディスクアレイのポート数およびランク数は必要に応じて適宜に定めることができる。

【0193】また第1発明の実施例にあつては、入出力デバイスとしてディスクアレイを例にとっているが、アレイ構成をもたない通常のディスク装置についてもそのまま適用することができる。更に第2発明にあつては、ホストコンピュータに対し2系統のキャッシュメモリを有するコントローラを設けた場合を例にとっているが、1系統のコントローラにキャッシュメモリを設けたものであったもよいことは勿論である。

【0194】更に本発明は実施例に示された数値による限定は受けない。

【0195】

【発明の効果】以上説明してきたように第1発明によれば、必要に応じてキャッシュメモリの信頼性を維持しながらの性能の向上、性能を維持しながらの信頼性の向上、および両者の中間的性能という柔軟な運用を選択的にとることができ、ディスクキャッシュを備えたシステム全体の運用効率を向上させることができる。

【0196】また第2発明にあつては、ディスクアレイを対象とした所謂RAID型の動作特性を活かしながらキャッシュ性能を最大限に発揮することができる。

【図面の簡単な説明】

【図1】本発明の原理説明図

【図2】本発明のハードウェア構成を示した実施例構成図

【図3】図2のコントローラのハードウェア構成を示した実施例構成図

【図4】本発明の第1発明の処理機能を示した説明図

【図5】図4の信頼性重視モードの処理動作を示したフローチャート

【図6】図4の信頼性重視モードにおけるライトバック処理を示したフローチャート

【図7】信頼性重視モードでのリード要求に対しローカル側でミスヒットとなった場合の動作を示した説明図

【図8】信頼性重視モードでのライト要求に対するキャッシュ書込動作の説明図

【図9】図4の性能重視モードの処理動作を示したフ

ローチャート

【図10】図4の性能重視モードにおけるライトバック処理を示したフローチャート

【図11】性能重視モードでのリード要求に対しローカル側でミスヒット、リモート側でヒットとなった場合の動作を示した説明図

【図12】性能重視モードでのライト要求に対するキャッシュ書込動作の説明図

【図13】図4の平均モードの処理動作を示したフローチャート

【図14】図4の平均モードにおけるライトバック処理を示したフローチャート

【図15】ディスクアレイを対象とした第2発明の処理機能を示した説明図

【図16】キャッシュメモリのブロック分けとホストブロックの関係を示した説明図

【図17】ホストブロックとデータブロックとの関係を示した説明図

【図18】図15のハッシュテーブルの説明図

【図19】図15のLRUテーブルの説明図

【図20】図15のキャッシュメモリの使用状態の一例を示した説明図

【図21】図15の空スペース管理テーブルの説明図

【図22】図15のディスクアレイに対するRAID0,1,3,5のディスク割当てを示す論理デバイスの説明図

【図23】RAID3におけるキャッシュブロック、ストライピング及びディスク格納状態を示した説明図

【図24】RAID5におけるキャッシュブロック、ストライピング、ディスク格納状態を示した説明図

【図25】RAID1におけるキャッシュブロックとディスク格納状態を示した説明図

【図26】RAID0におけるキャッシュブロックとディスク格納状態を示した説明図

【図27】ステージング処理に伴うプリフェッチ動作を示した説明図

【図28】プリフェッチのステージングを含む図15のリード処理を示したフローチャート

【図29】大量シーケンシャルデータのステージング処理を示したフローチャート

【図30】パリティを含むステージング処理を示したフローチャート

【図31】パリティを含まない場合と含む場合のステージングに対応するライトバック処理の相違を示した説明図

【図32】LRUテーブルにおけるライトバック対象範囲の説明図

【図33】ライトバック対象範囲におけるキャッシュブロック並び替えとライトバック動作を示した説明図

【図34】ライトバック完了後のLRUテーブルの説明

図

【図35】図27のライト処理の詳細を示したフローチャート

【図36】RAID5におけるライトバック動作の説明図

【図37】RAID5において不足ブロックをキャッシュブロックにステージングするライトバック動作の説明図

【図38】ライト完了後に行うライトバック処理を示したフローチャート

【図39】キャッシュブロックサイズの求める統計処理のフローチャート

【図40】最適なキャッシュブロックサイズの決め方の一例を示した説明図

【図41】揮発性メモリを用いた従来のディスクキャッシュ装置の説明図

【図42】図41の従来装置のキャッシュヒット時のリード処理を示した説明図

【図43】図41の従来装置のキャッシュミスヒット時のリード処理を示した説明図

【図44】図41の従来装置のライト処理を示した説明図

【図45】揮発性メモリと不揮発性メモリを混在させた従来装置の説明図

【図46】図45の従来装置のキャッシュミスヒット時のリード処理を示した説明図

【図47】図45の従来装置のライト処理を示した説明図

【符号の説明】

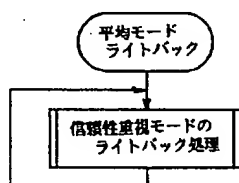
10：ホストコンピュータ（上位装置）
12，12-1，12-2：コントローラ（ディスク制御手段）
14-1，14-2：チャネル装置
16：チャネルインタフェース（SCSI）
18，18-1，18-2：共用バス（共用バス手段）
20：ブリッジ回路

22-1，22-2：サブコントローラ
24-1～24-6，26-1～26-6：アダプタ
28：ディスクアレイ
30，30-00～30-35：ディスク装置
32：CPU
34：ROM
36：DRAM
38：上位インタフェース部
40：SCSI回路部
42：バスインタフェース部
44：内部バス
45：ライトバック処理部
46：キャッシュ制御部
46-1：リードキャッシュ制御部
46-2：ライトキャッシュ制御部
48：キャッシュメモリ
48-1，48-2：不揮発性キャッシュメモリ
50：キャッシュ管理テーブル
52：ディスクアレイ制御部
54，54-1～54-4：キャッシュブロック
56：ホストブロック
58：デバイスブロック（1セクタ）
60：ハッシュテーブル
62：LRUテーブル
64：空スペース管理テーブル
66：ハッシュエントリ
68：登録データ
70，70-1，70-2：動作モード設定部
72，72-1，72-2：信頼性重視モードキャッシュ制御部
74，74-1，74-2：性能重視モードキャッシュ制御部
75：ライトバック対象範囲
76，76-1，76-2：平均モードキャッシュ制御部

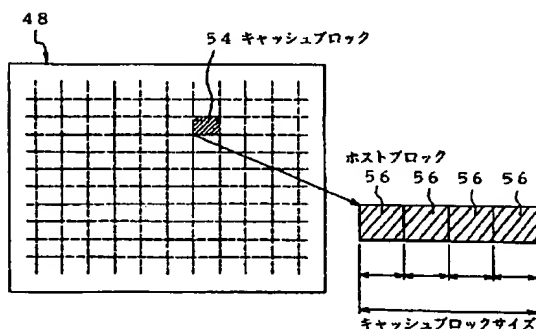
【図14】

【図16】

図4の平均モードにおけるライトバック処理を示したフローチャート



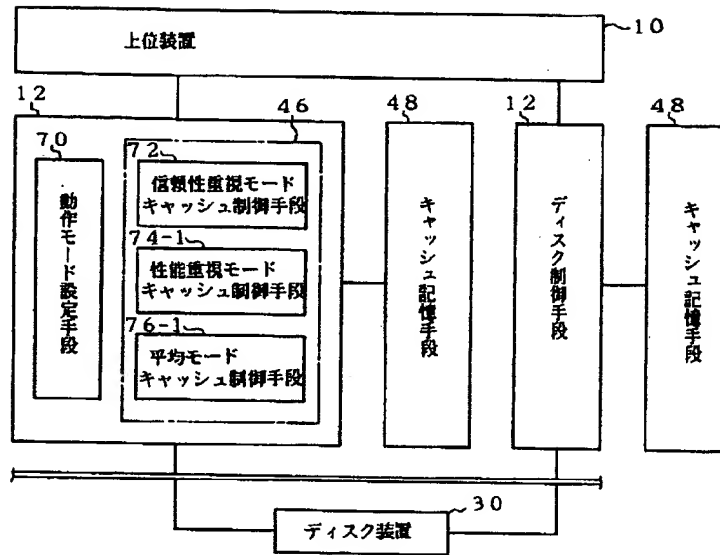
キャッシュメモリのブロック分けとホストブロックの関係を示した説明図



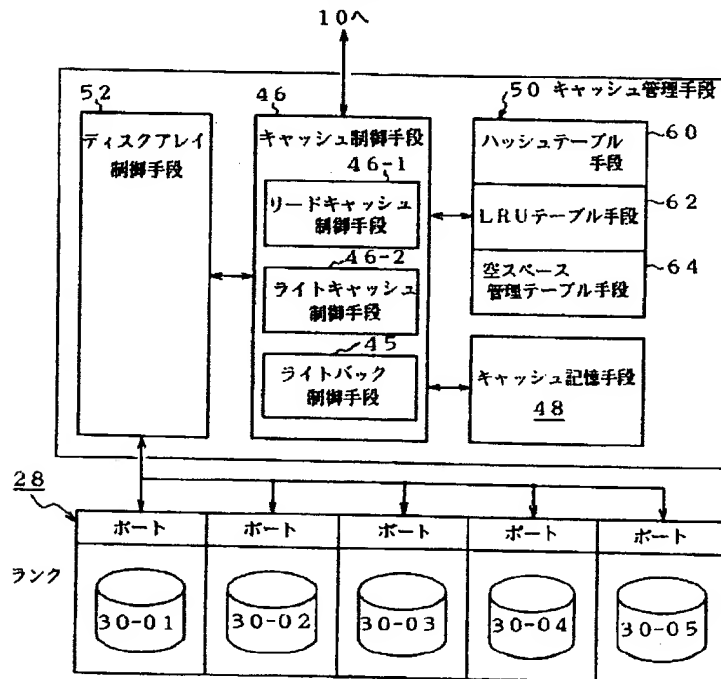
【図1】

本発明の原理説明図

(A) 第1発明

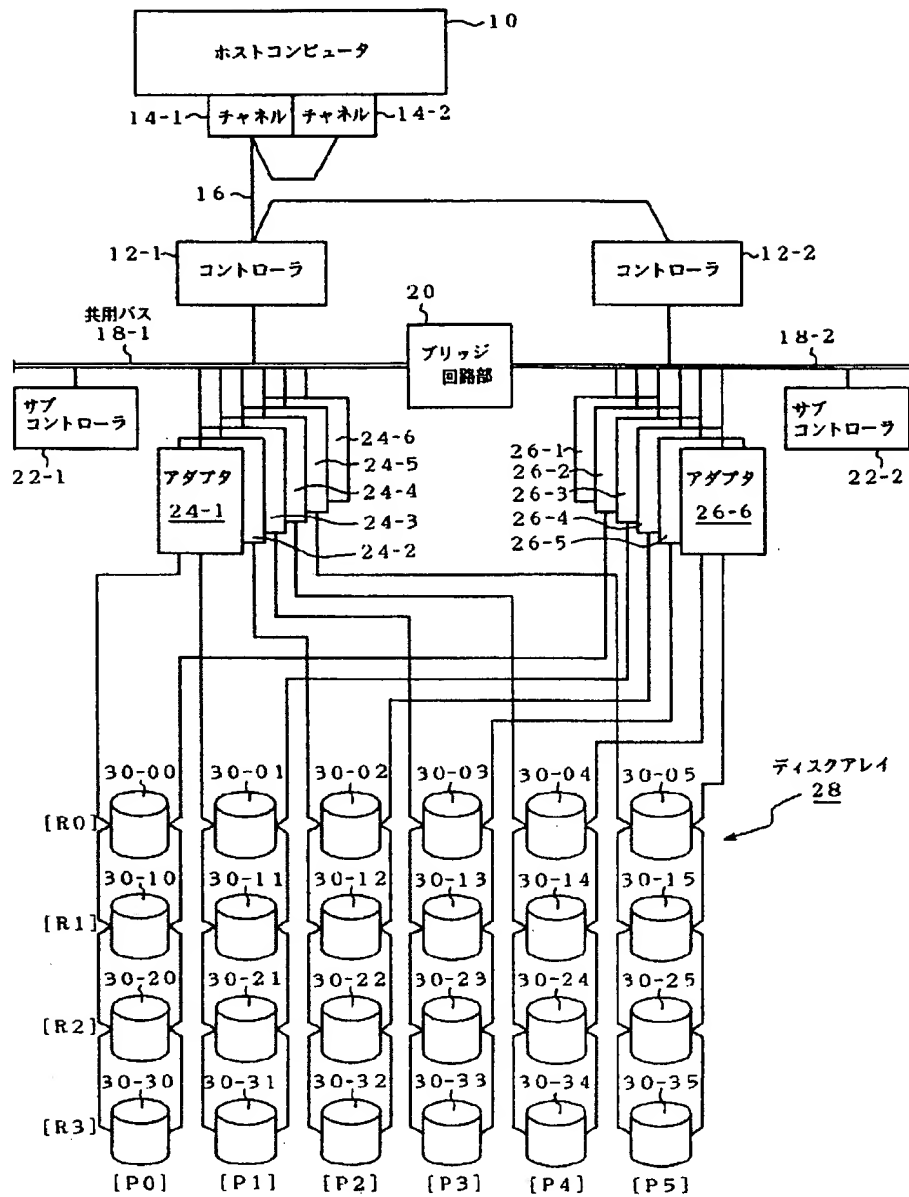


(B) 第2発明



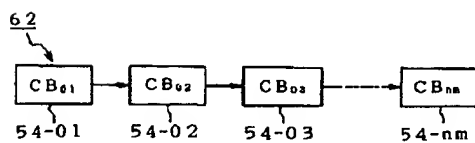
【図2】

本発明のハードウェア構成を示した実施例構成図



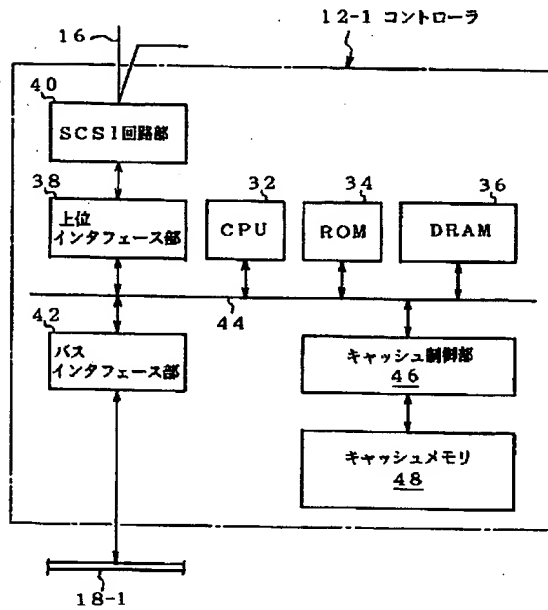
【図19】

図15のLRUテーブルの説明図



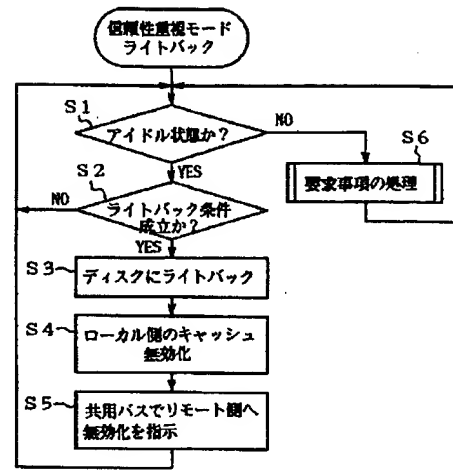
【図3】

図2のコントローラのハードウェア構成を示した実施例構成図



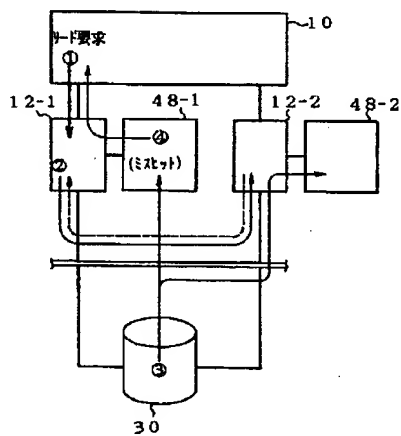
【図6】

図4の信頼性重視モードにおけるライトバック処理を示したフローチャート



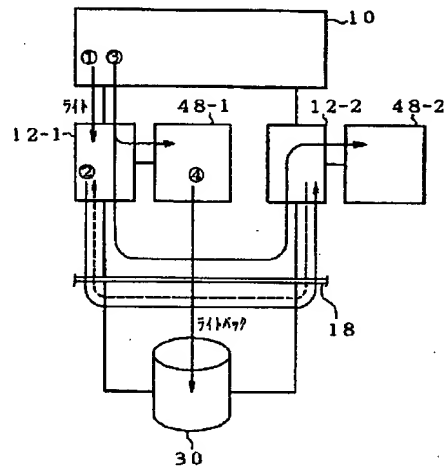
【図7】

信頼性重視モードでのリード要求に対しローカル側でミスヒットとなった場合の動作を示した説明図

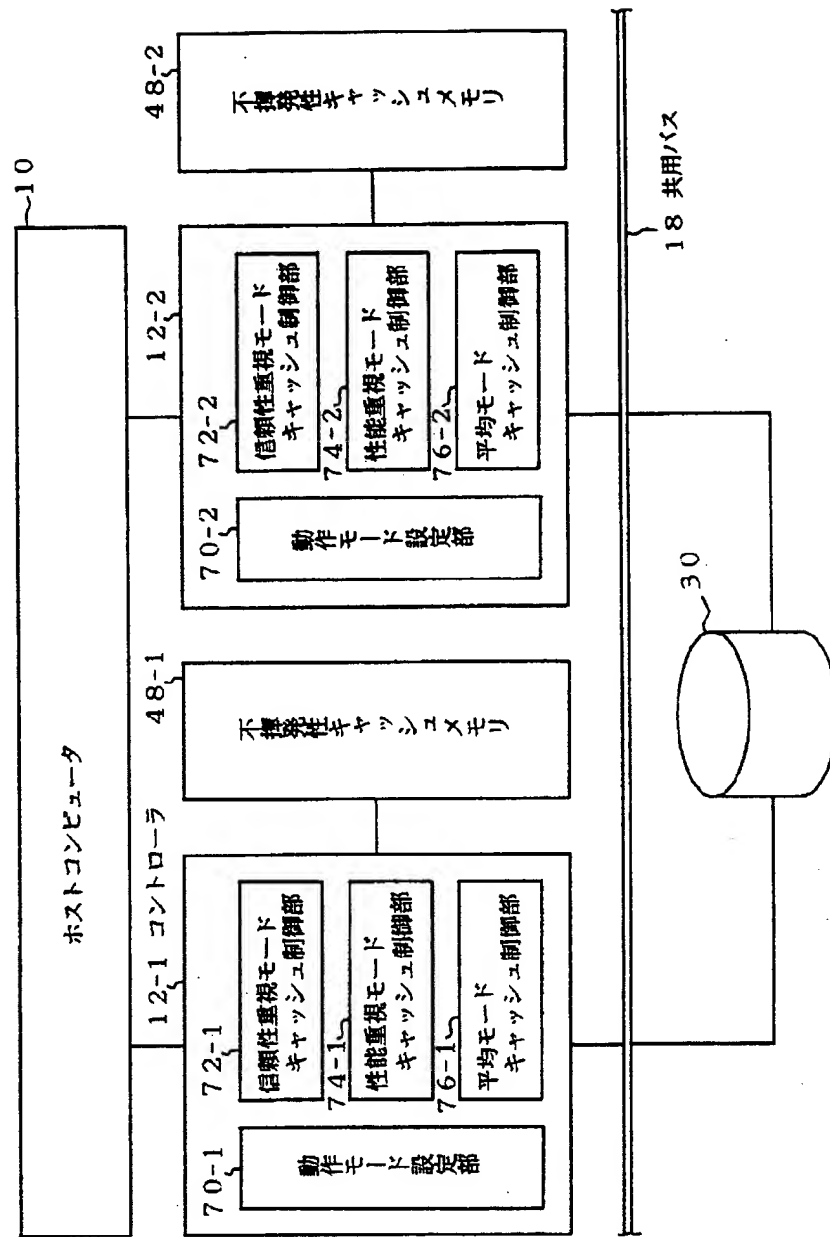


【図8】

信頼性重視モードでのライト要求に対するキャッシュ書込動作の説明図



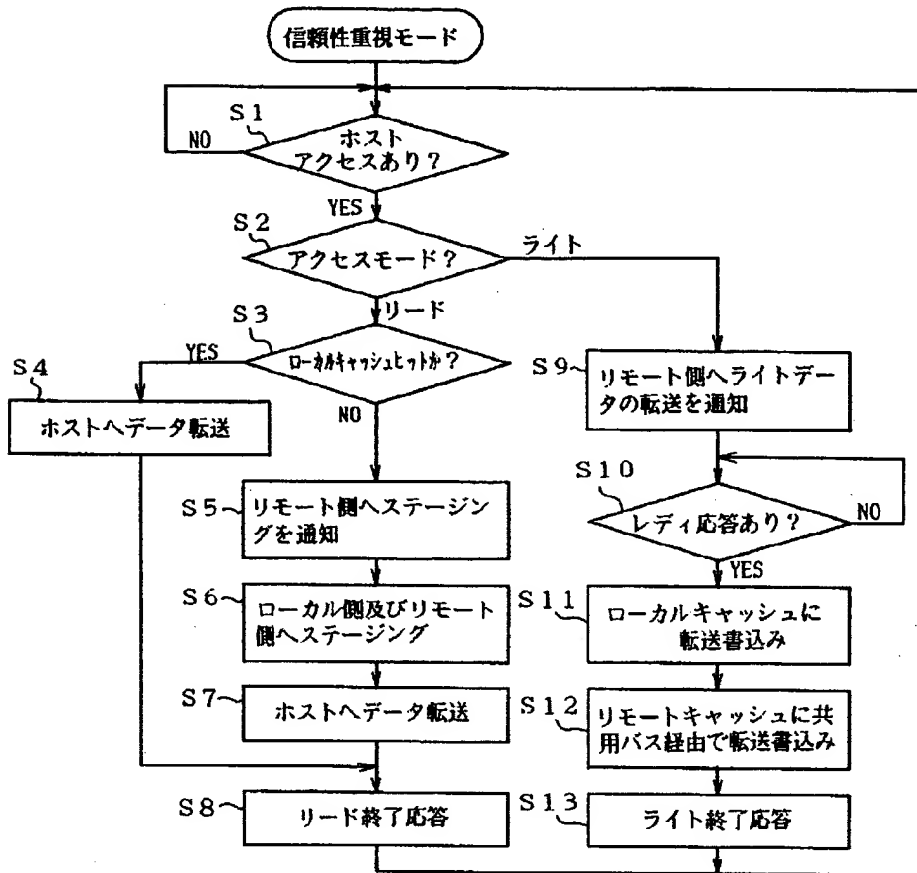
本発明の第1発明の処理機能を示した説明図



【図4】

【図5】

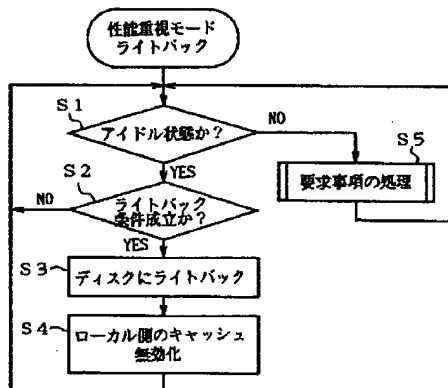
図4の信頼性重視モードの処理動作を示したフローチャート



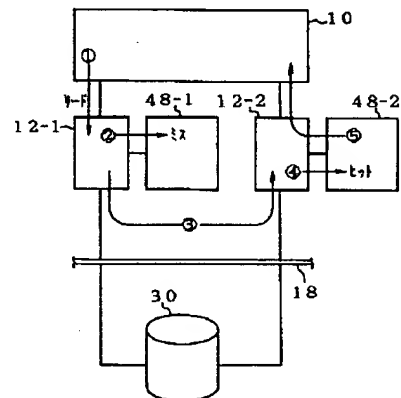
【図10】

【図11】

図4の性能重視モードにおけるライトバック処理を示したフローチャート

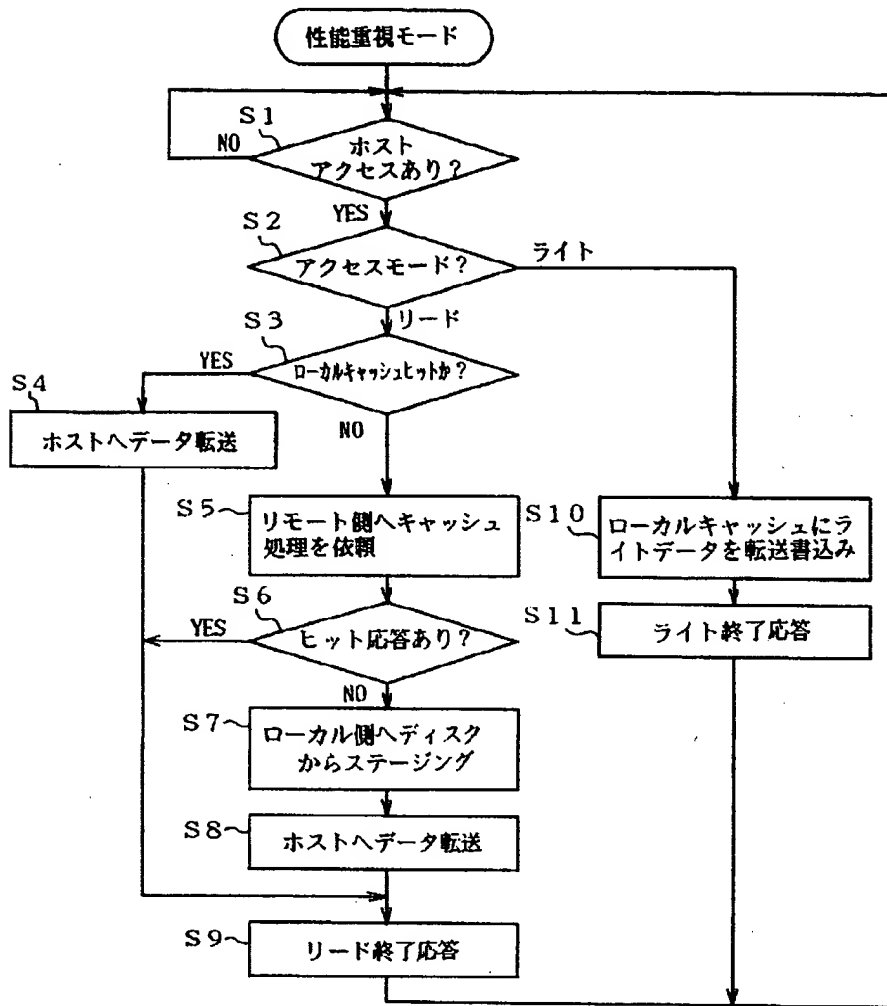


性能重視モードでのリード要求に対しローカル側でミスヒット、リモート側でヒットとなった場合の動作を示した説明図



【図9】

図4の性能重視モードの処理動作を示したフローチャート



【図21】

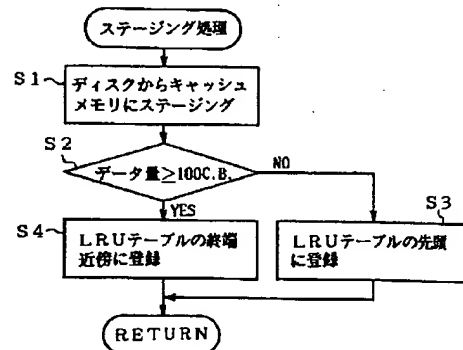
【図29】

図15の空スペース管理テーブルの説明図

64

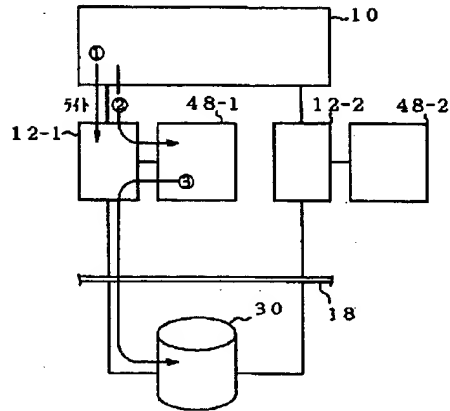
エントリ	先頭ブロック	内 容
1	CB0203	キャッシュブロック1個分の空スペース有り
	CB0004	
2	CB0102	キャッシュブロック2個分の空スペース有り
3	CB00n-2	キャッシュブロック3個分の空スペース有り
⋮	⋮	⋮
N	CB00n-1	キャッシュブロックN個分の空スペース有り

大量シーケンシャルデータのステーjing処理を示したフローチャート



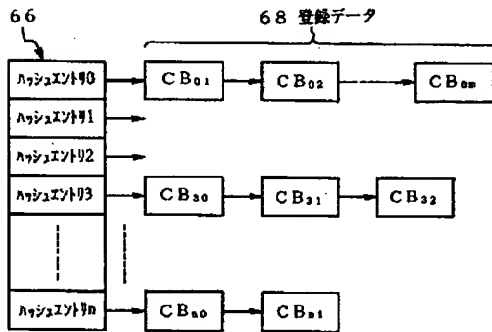
【図12】

性能重視モードでのライト要求に対するキャッシュ書込動作の説明図



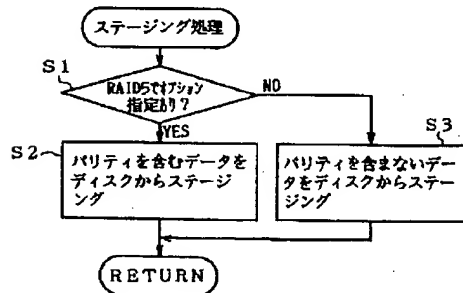
【図18】

図15のハッシュテーブルの説明図



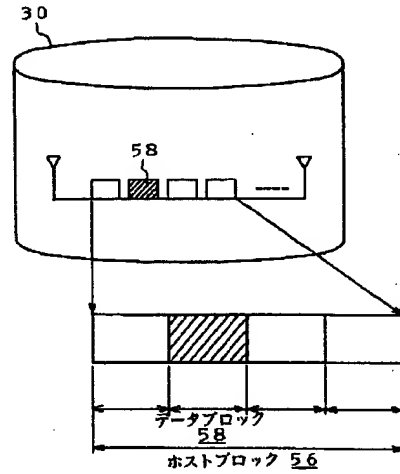
【図30】

パリティを含むステージング処理を示したフローチャート



【図17】

ホストブロックとデータブロックとの関係を示した説明図



【図20】

図15のキャッシュメモリの使用状態の一例を示した説明図

		48							
		00	01	02	03	04	05	06	07
00	R	R	W	R	R	R	R	R	R
01	R	W	W	W	W	W	W	W	W
02	R	空	空	R	R	R	R	R	R
03	R	空	R	R	R	R	R	R	R
04	空	R	R	R	R	R	R	R	R
...
n-3	R	R	R	R	W	W	W	W	W
n-2	空	空	空	R	R	R	R	R	R
n-1	空	空	空	空	空	空	空	空	空
n	空	空	空	空	空	空	空	空	空

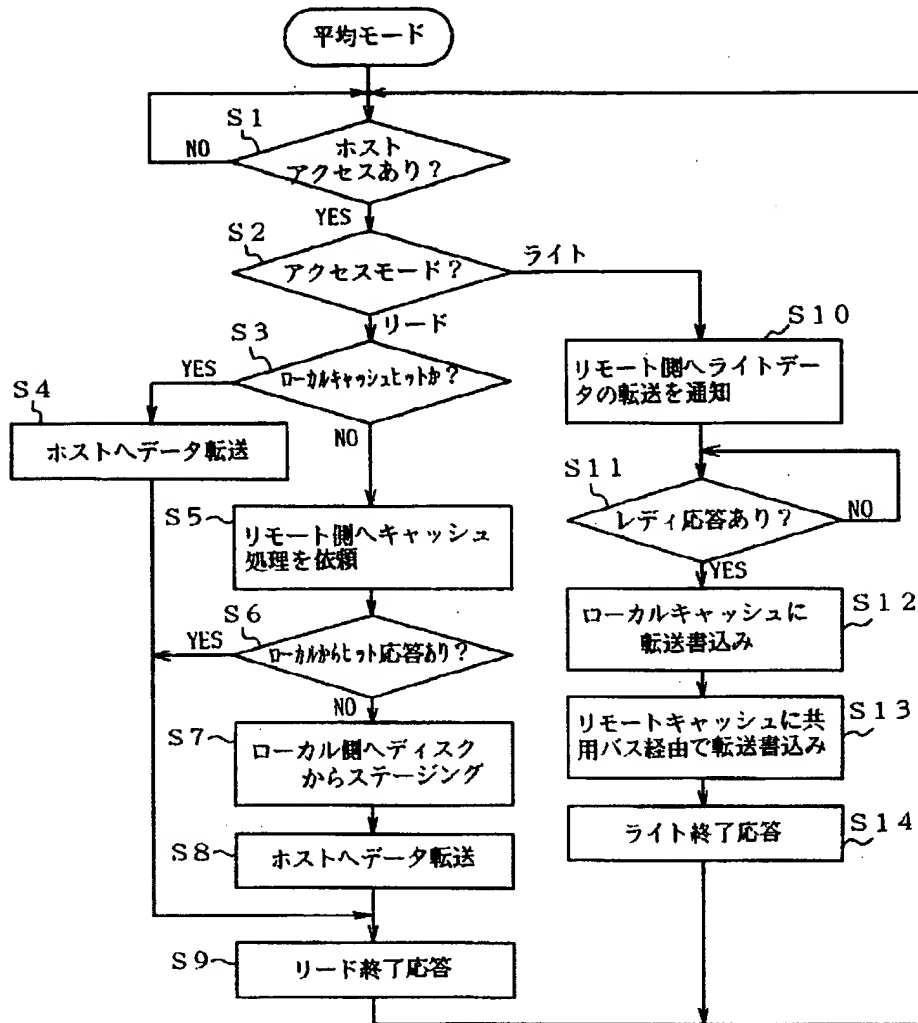
R: 有効データの存在するキャッシュブロック

W: ダーティデータの存在するキャッシュブロック

空: 空キャッシュブロック

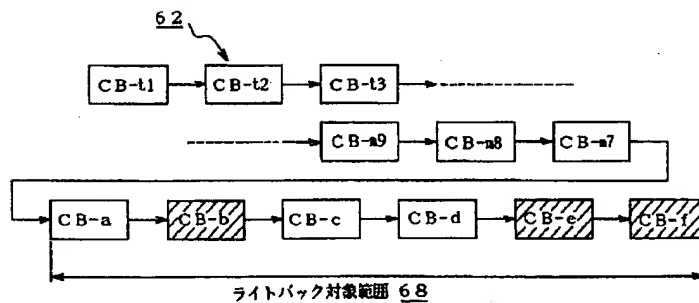
【図13】

図4の平均モードの処理動作を示したフローチャート



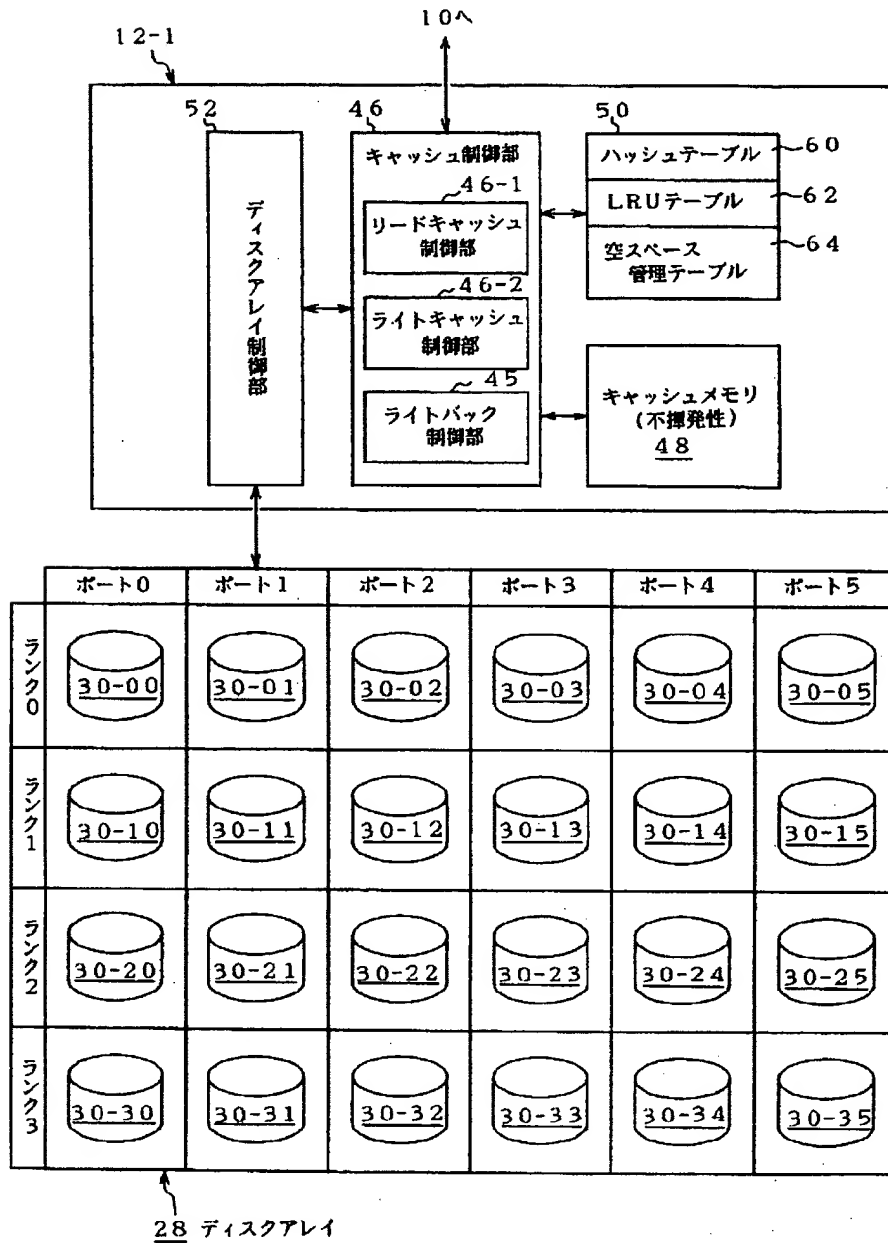
【図32】

LRUテーブルにおけるライトバック対象範囲の説明図



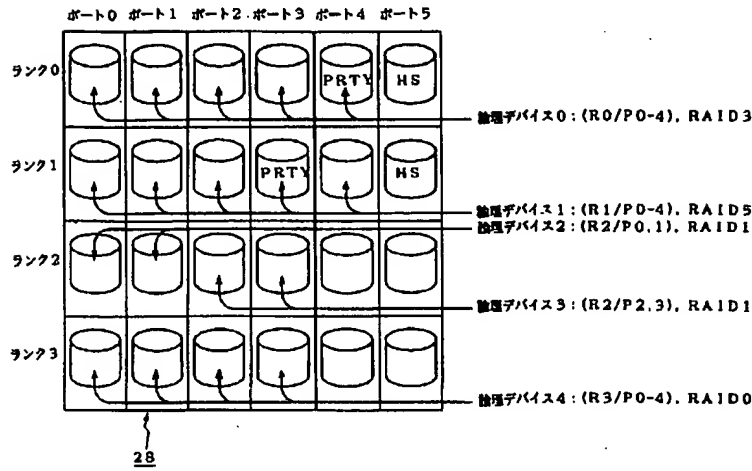
【図15】

ディスクアレイを対象とした第2発明の処理機能を示した説明図



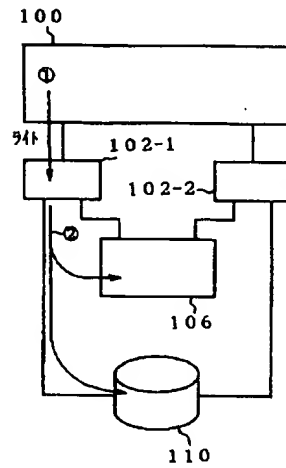
【図 2 2】

図15のディスクアレイに対するRAID0, 1, 3, 5のディスク割当てを示す論理デバイスの説明図



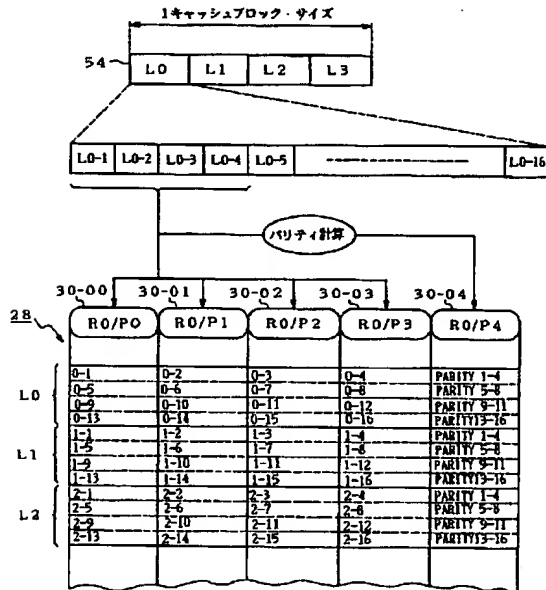
【図 4 4】

図41の従来装置のライト処理を示した説明図



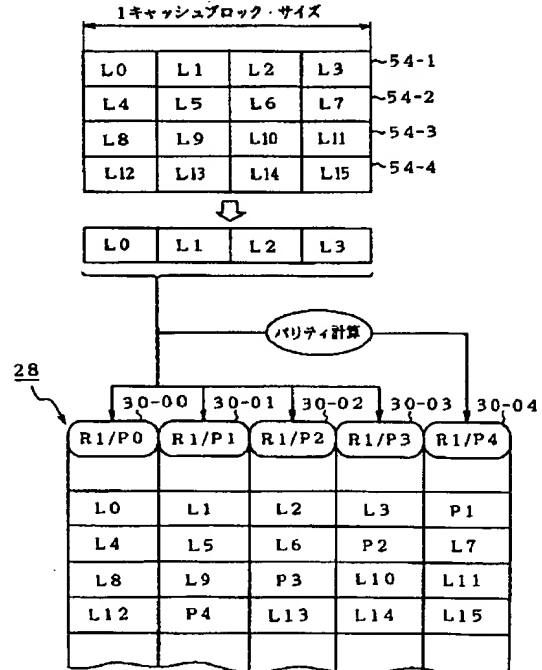
【図 2 3】

RAID3におけるキャッシュブロック、ストライピング及びディスク格納状態を示した説明図



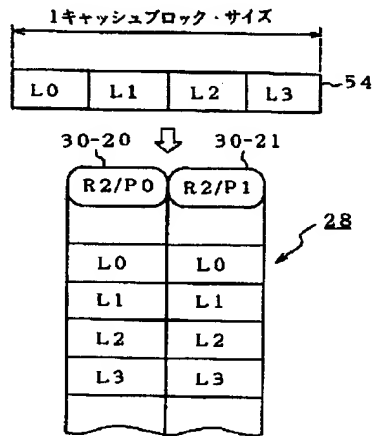
【図 2 4】

RAID5におけるキャッシュブロック、ストライピング、ディスク格納状態を示した説明図



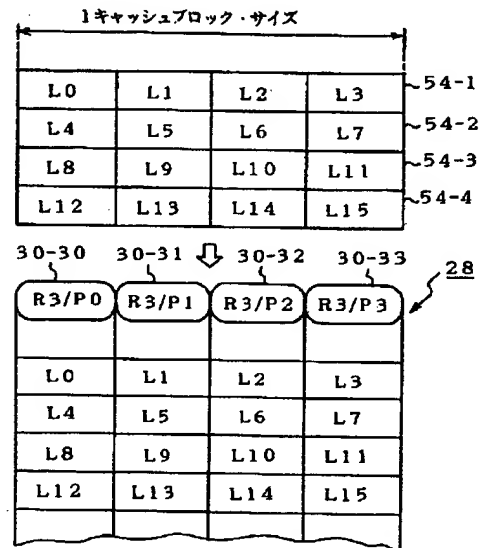
【図25】

RAID1におけるキャッシュブロックとディスク格納状態を示した説明図



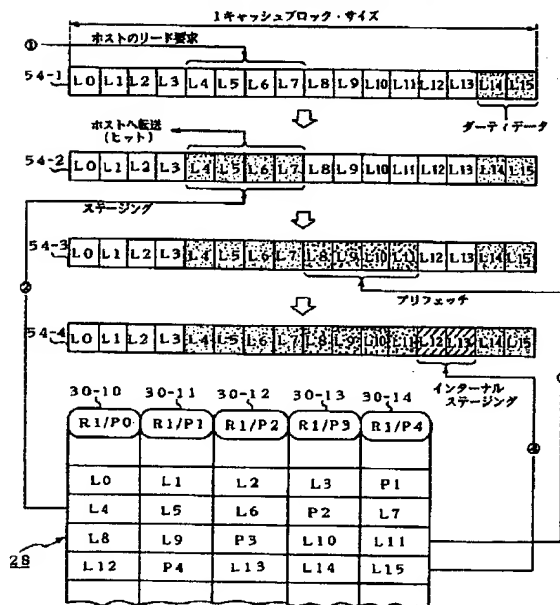
【図26】

RAID0におけるキャッシュブロックとディスク格納状態を示した説明図



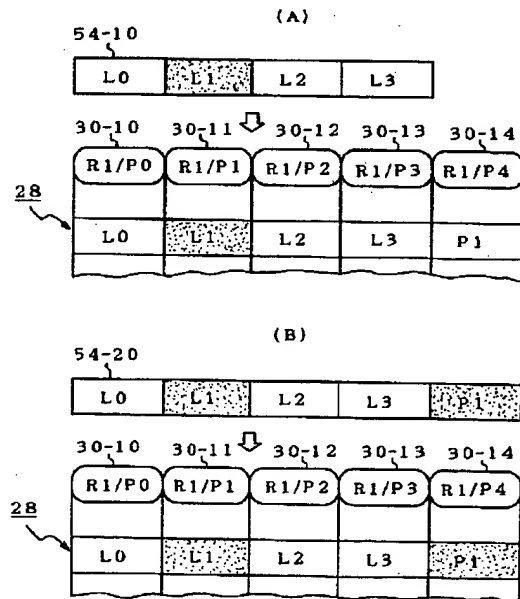
【図27】

ステージング処理に伴うプリフェッチ動作を示した説明図



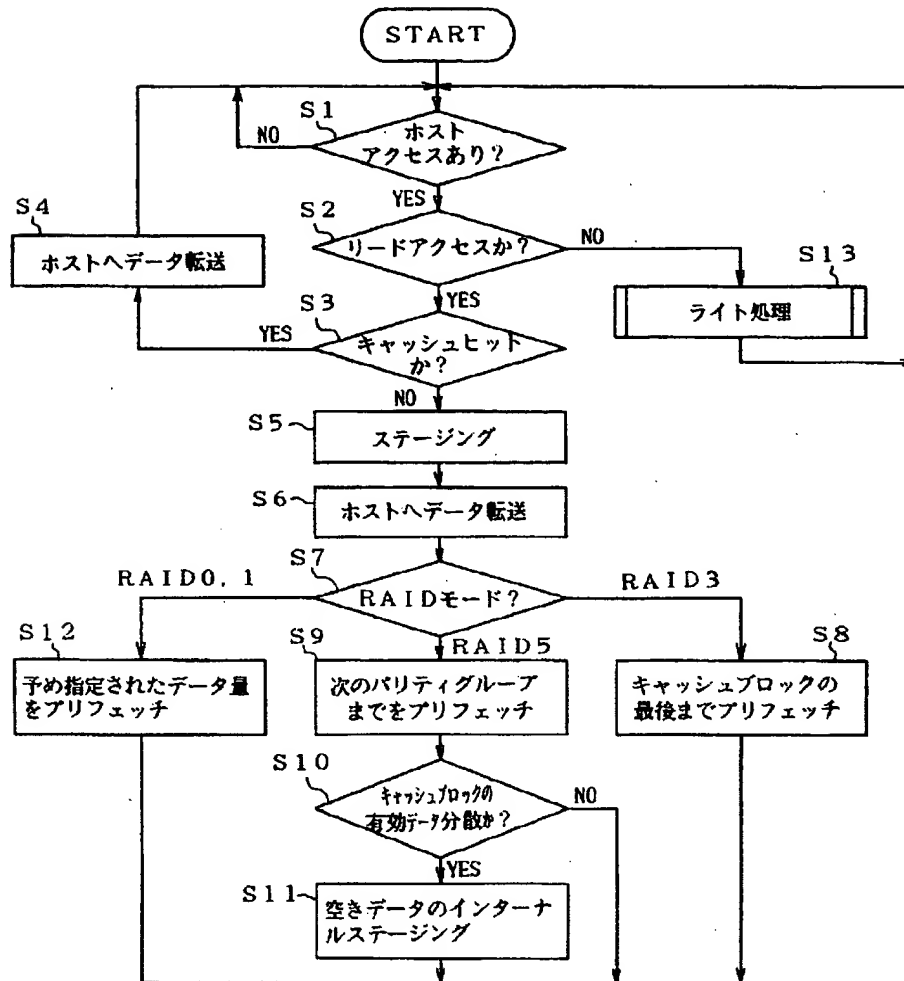
【図31】

パリティを含まない場合と含む場合のステージングに対応するライトバック処理の相違を示した説明図



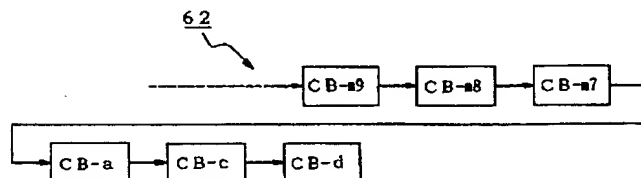
【図28】

プリフェッチのステージングを含む図15のリード処理を示したフローチャート



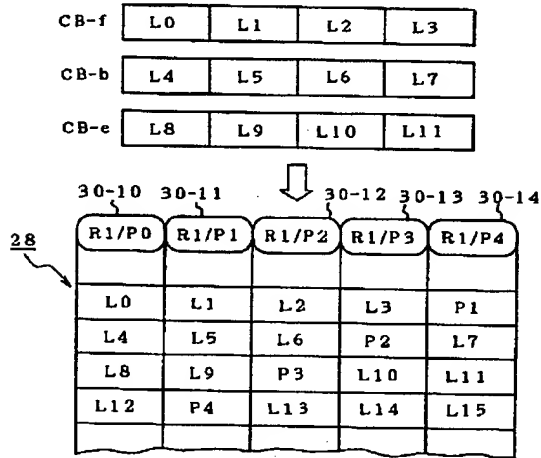
【図34】

ライトバック完了後のLRUテーブルの説明図



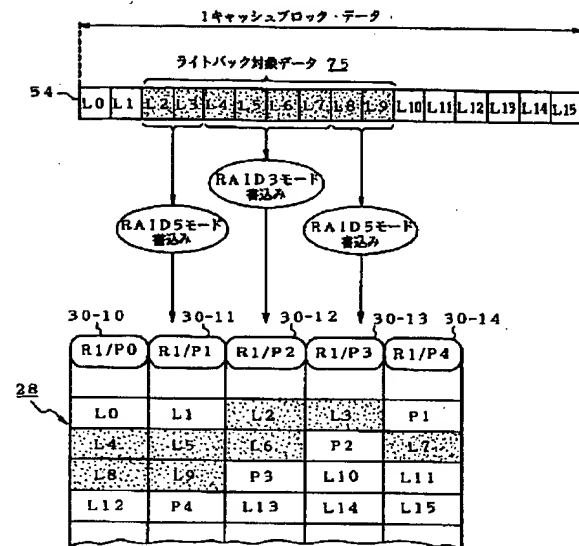
【図33】

ライトバック対象範囲におけるキャッシュブロック並び替えとライトバック動作を示した説明図



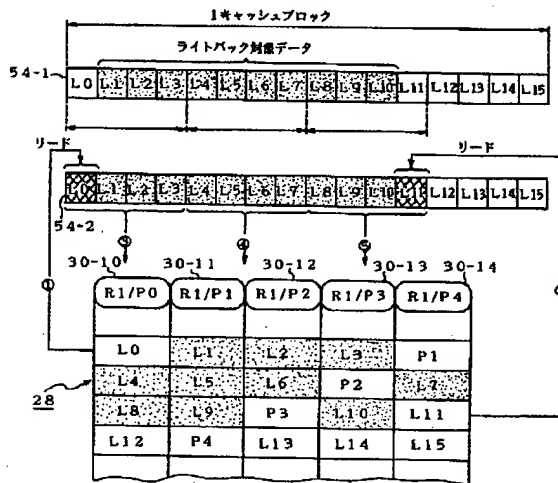
【図36】

RAID5におけるライトバック動作の説明図



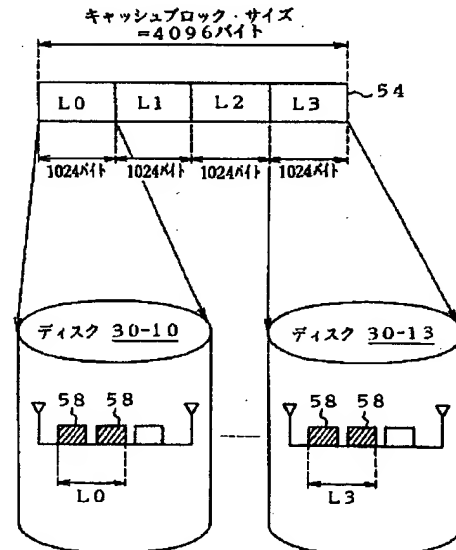
【図37】

RAID5において不足ブロックをキャッシュブロックにステージングするライトバック動作の説明図



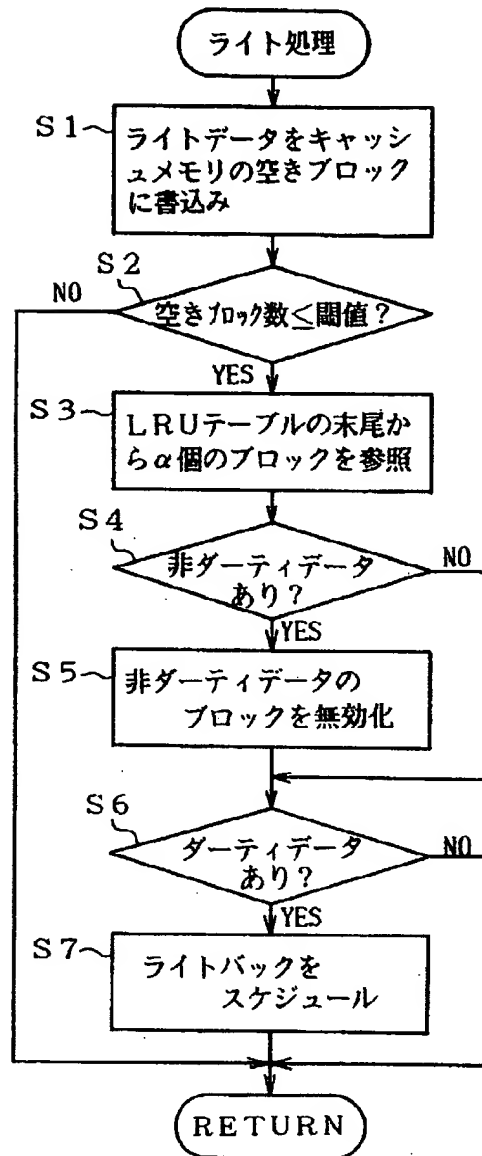
【図40】

最適なキャッシュブロックサイズの決め方の一例を示した説明図



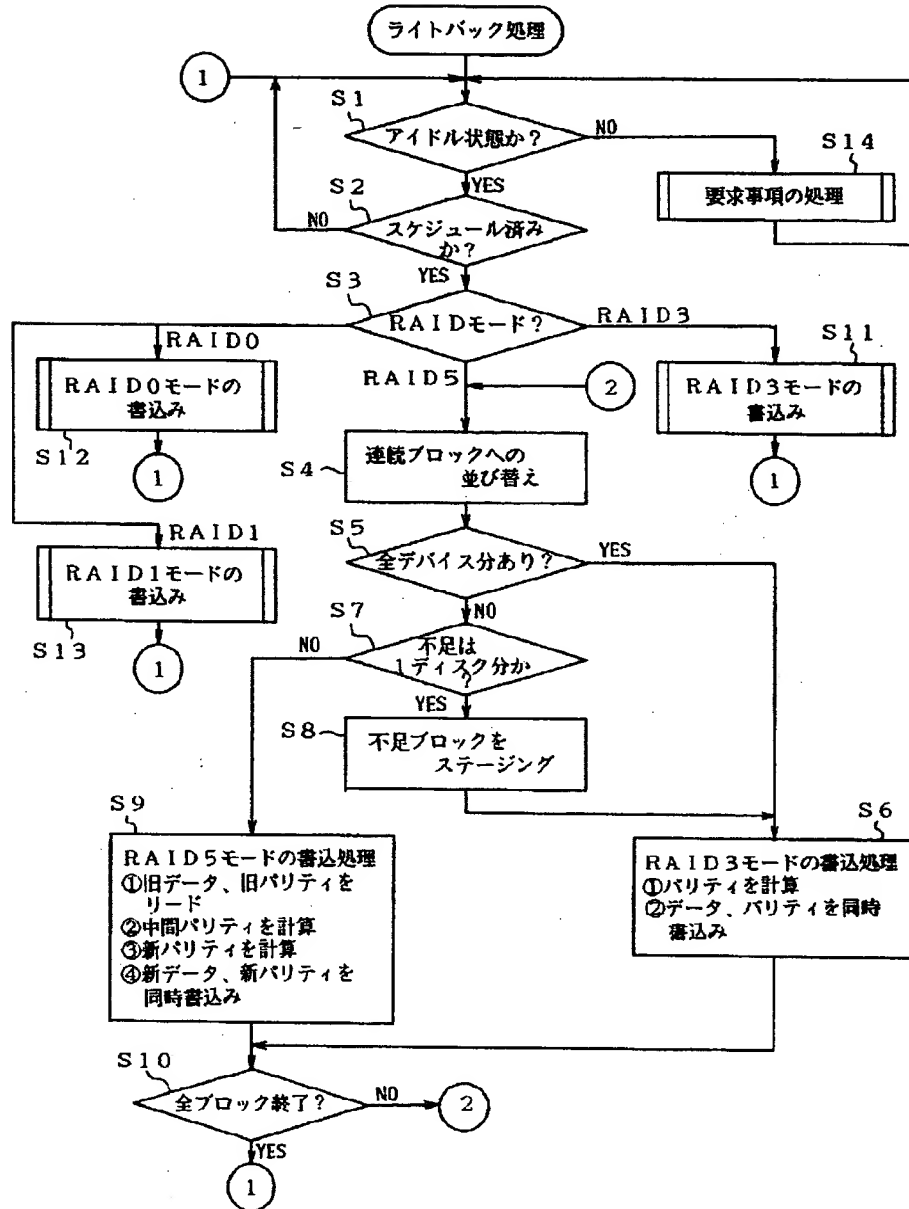
【図35】

図27のライト処理の詳細を示したフローチャート



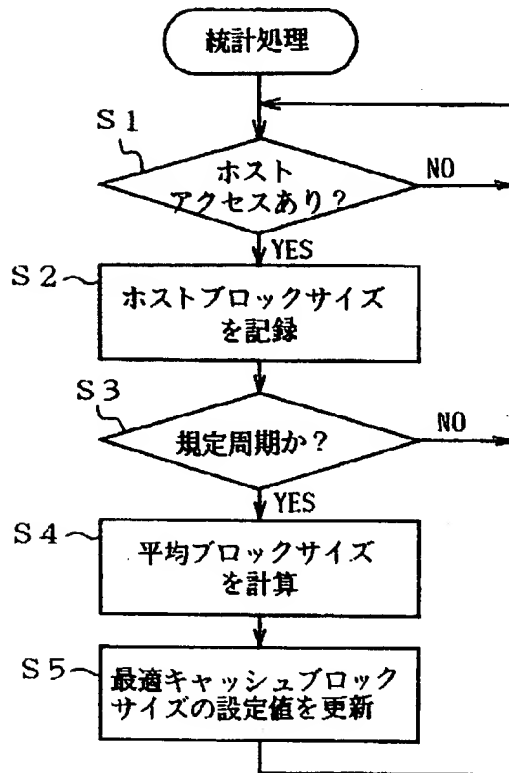
【図38】

ライト完了後に行うライトバック処理を示したフローチャート



【図 3 9】

キャッシュブロックサイズの求める統計処理のフローチャート

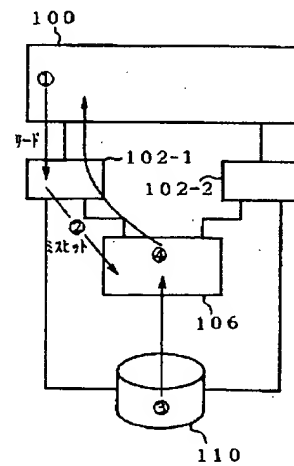
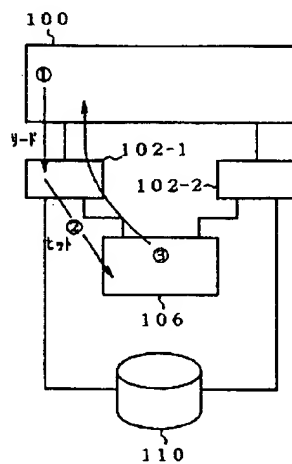


【図 4 2】

【図 4 3】

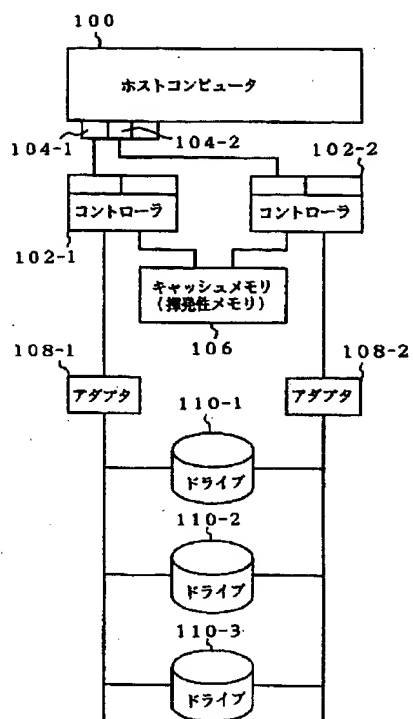
図 41 の従来装置のキャッシュミスヒット時のリード処理を示した説明図

図 41 の従来装置のキャッシュヒット時のリード処理を示した説明図



【図 4 1】

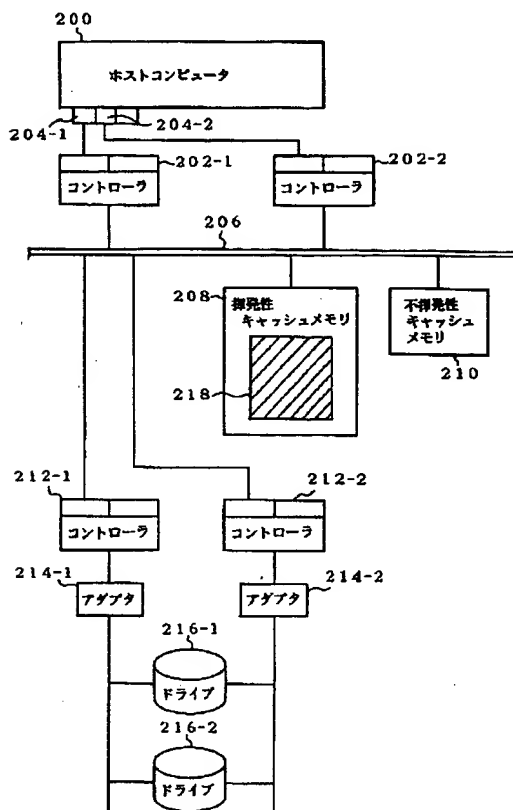
揮発性メモリを用いた従来のディスクキャッシュ装置の説明図



【図 4 6】

【図 4 5】

揮発性メモリと不揮発性メモリを混在させた従来装置の説明図



【図 4 7】

図 4 5 の従来装置のキャッシュミスヒット時のリード処理を示した説明図

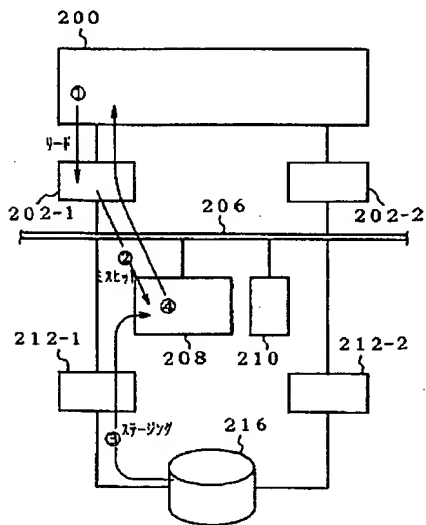


図 4 5 の従来装置のライト処理を示した説明図

